HP Computer Systems

Statistical Library

for the HP 9826 and 9836 Computers

HEWLETT PACKARD

HEWLETT
PACKARD

Warranty Statement

Hewlett-Packard makes no expressed or implied warranty of any kind, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose, with regard to the program material contained herein. Hewlett-Packard shall not be liable for incidental or consequential damages in connection with, or arising out of, the furnishing, performance or use of this program material.

HP warrants that its software and firmware designated by HP for use with a CPU will execute its programming instructions when properly installed on that CPU. HP does not warrant that the operation of the CPU, software, or firmware will be uninterrupted or error free.

Use of this manual and flexible disc(s) supplied for this pack is restricted to this product only. Additional copies of the programs can be made for security and back-up purposes only. Resale of the programs in their present form or with alterations, is expressly prohibited.

Restricted Rights Legend

Use, duplication, or disclosure by the Government is subject to restrictions as set forth in paragraph (b)(3)(B) of the Rights in Technical Data and Software clause in DAR 7-104.9(a).

Statistical Library

for the HP 9826 and 9836 Computers

Manual Part No. 98820-13111

Disc Part Numbers

Basic Statistics	98820-13114
General Statistics	98820-13115
Statistical Graphics I	98820-13116
Statistical Graphics II	98820-13117
Regression Analysis	98820-13118
Analysis of Variance I	98820-13124
Analysis of Variance II	98820-13125
Principle Components and Factor Analysis	98820-13126
Monte Carlo Routines	98820-13127
Monte Carlo Tests	98820-13128

Important

The flexible disc containing the programs is very reliable, but being a mechanical device, is subject to wear over a period of time. To avoid having to purchase a replacement medium, we recommend that you immediately duplicate the contents of the disc onto a permanent backup disc. You should also keep backup copies of your important programs and data on a separate medium to minimize the risk of permanent loss.

Hewlett-Packard Desktop Computer Division
3404 East Harmony Road, Fort Collins, Colorado 80525

Copyright by Hewlett-Packard Company 1982

Printing History

New editions of this manual will incorporate all material updated since the previous edition. Update packages may be issued between editions and contain replacement and additional pages to be merged into the manual by the user. Each updated page will be indicated by a revision date at the bottom of the page. A vertical bar in the margin indicates the changes on each page. Note that pages which are rearranged due to changes on a previous page are not considered revised.

The manual printing date and part number indicate its current edition. The printing date changes when a new edition is printed. (Minor corrections and updates which are incorporated at reprint do not cause the date to change.) The manual part number changes when extensive technical changes are incorporated.

July 1982...First Edition

Table of Contents

Table 1
Statistical Distributions

Table Values and Right-Tail Probabilities

Continuous	Discrete
1. Normal	1. Binomial
2. Two-paremeter gamma	2. Negative Binomial
3. Central F	3. Poisson
4. Beta	4. Hypergeometric
5. Student's T	5. Gamma Function
6. Weibull	6. Beta Function
7. Chi-square	7. Single Term Binomial
8. Laplace	8. Single Term Negative Binomial
9. Logistic	9. Single Term Poisson
	10. Single Term Hypergeometric

Commentary

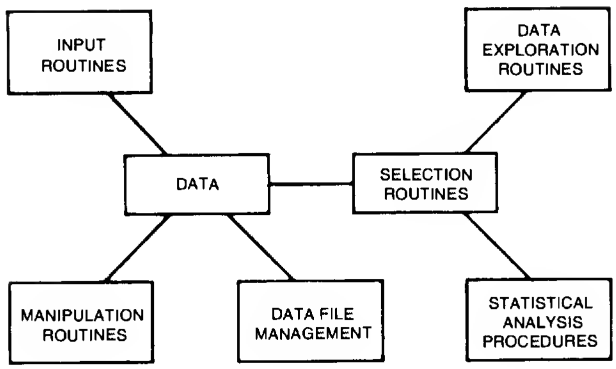
The Stat Library, which we have developed for Hewlett-Packard, is an integrated package developed specifically for the HP desktop computers. We set as our objective in preparing this library to develop an integrated system which provides the user with a flexible collection of routines for **data manipulation**, **exploration**, and **analyses.** The package uses a common front end, which provides for considerable flexibility in data handling. The Basic Statistics and Data Manipulation (BSDM) front end has been updated and enhanced for inclusion with this library. The programs are interactive in operation using the CRT display to list a "menu" of options at appropriate times. The group of special function keys are used only with the BSDM routines to connect the user directly with a specific operation. The statistical analyses range from the very elementary summary statistics to complicated routines for principal com-ponents and factor analysis.

The figure on the next page is a diagram showing the essential organizational structure of the Stat Library. Notice that there are six major segments in the Stat Library which operate on the data: Input Routines, Manipulation Routines, Data File Management Routines, Selection Routines, Data Exploration Routines, and Statistical Analysis Procedures.

This library has evolved out of our ten years' experience in developing software for desktop computers. We are currently using these routines in our Statistical Laboratory. We hope you will find them useful.

Thomas J. Boardman, Ph.D.
Professor-In-Charge
Statistical Laboratory
Colorado State University
Fort Collins, CO 80523

HP Stat Library
Integrated Statistical Routines



Operation (Key Words)	Description	Subprogram Package Containing Routine
Input Routines		BSDM
Keyboard	Direct numeric input by the user.	
Mass Storage	Of data previously stored on one of several mass storage devices.	
Graphics Input	Using the Graphics Tablet	
Other	User supplied routines	
Manipulation Routines		
Sort	Sorting data on one or two variables.	
Join	Joining two data sets either by adding variables or observations to existing set.	
Rename	Change variable label, subfile name, or project title.	
Subfile	Several methods to specify or create subfiles (groups within your data set).	
Recode	Method to recode variable values into another variable.	
Edit	To correct, add, or delete observations or variables.	
Transformation	By algebraic routines including user supplied function. To assign missing values. To create new variables by using ranks, subfile codes, sequence numbers, standardized scores, or lagged variables.	
Data Recovery	A backup data file may be accessed if necessary.	

(Continued)

Data File Management Routines		BSDM
Store	Save data set on user file.	
Store Subfile(s)	Save particular subfile on a user file.	
Store Variables	Save particular variables on a user file.	
Direct	Obtaining a catalog or directory of data file(s).	
Purge	Eliminate selected data files.	
Selection Routines		BSDM
By Subfiles	To choose a portion of the data for further analyses.	
Exclude Missing Values	Always excluded from analyses and data exploration routines.	
Select	To choose a portion of the data set for further processing on the basis of values from one or two variables. The values selected are shown on the CRT and the data set is reduced down to the selected data set size.	
Data Exploration Routines		
Selected Listing	Several ways are available to list all or a portion of the data set.	BSDM
Scan	Same as Select (above) except that data set is not reduced.	BSDM
Summary Statistics	Many basic statistics such as mean, median, standard deviation, etc., on all or a portion of the data set.	BSDM
Graphics Displays	Eight common statistical graphics for studying data sets such as normal probability plots and semi-log plots.	Stat Graphics
Frequencies	Under development for future addition to library.	
Cross Tabulation	Under development for future addtion to library.	

(Continued)

Statistical Analysis Procedures

General Parametric Methods	Common one, two-independent, and two-paired sample inferential procedures. Also one way analysis of variance.	General Statistics
General Nonparametric Method	Common one, two-independent, and two-paired sample nonparametric inferential procedures. Also the Kruskal Wallis test for 3 or more independent samples.	General Statistics
Regression Analyses Polynomial Multiple Linear Regression Stepwise	Selection procedures including the stepwise, forward, backward, and manual routines.	Regression Analysis
Nonlinear	From user supplied functions using the Marquardt Compromise algorithm.	
Standard Nonlinear	Several common nonlinear models are available for use on your data set.	
Analysis of Variance (AOV)		Analysis of Variance
One Way	One way AOV procedure.	
One Way Covariance	One way analysis of covariance procedure.	
Two Way Unbalanced	The AOV procedure for two way factorials which are unbalanced.	
Factorial	AOV procedure for up to 5 factors with balanced data.	
Split Plot	AOV methods for several types of split plot designs with up to 4 factors.	
Nested	AOV methods for completely or partially balanced nested designs.	
Principal Components and Factor Analysis	Common multivariable dimension reduction procedures. Extensive use of graphics.	Principal Components and Factor Analysis
(Others)	In the future.	

Basic Statistics and Data Manipulation

General Information

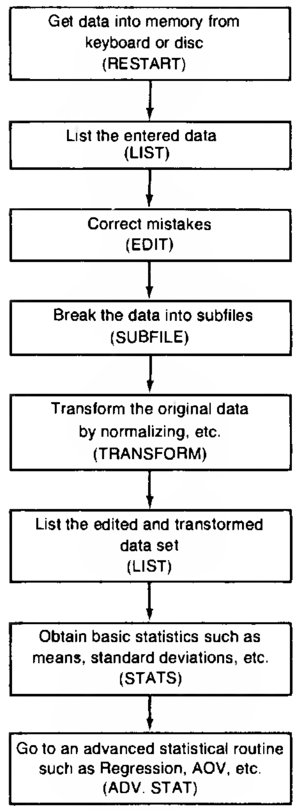
Description

This set of programs allows you to create a statistical data base which can be accessed by other Hewlett Packard statistical routines. It alleviates the need to key in data each time a new statistical procedure is used.

The capabilities of this set of programs include data entry and several manipulative data operations. A wide variety of summary statistics may be obtained. In addition, the programs have many ease-of-use features – the human interface is a major concern in designing the programs. Specific capabilities follow.

Data Entry:	Keyboard Magnetic media (flexible discs) Graphics tablet Other input devices (paper tape, etc.)
Data Manipulation:	Edit incorrect/incomplete data sets Transform – both algebraic and non-algebraic Assign codes to intervals of data Sort Divide data set into subfiles Join two data sets Select portions of the data
Summary Statistics:	Basic statistics (mean, standard deviation, etc.) Correlation matrix Order statistics (max, min, median, etc.)
Other Features:	Error detection Easy error correction Variables can be named Data can be stored for future reference Data can be listed Data can be scanned for specified qualities A backup file of the data can be recalled Printer unit can be changed Missing data values can be assigned

Typical Program Flow



Special Considerations

Data Matrix Configuration

The data matrix incorporated in this program should be thought of as a p-by-n array whose columns correspond to observations and whose rows correspond to variables as shown below.

		OBSERVATIONS				
		O_1	O_2	O_3	...	O_n
	V_1					
	V_2					
VARIABLES	V_3					
	.					
	.					
	.					
	V_p					

Subfiles may be created, in which case the structure becomes only slightly more complex as shown below.

		OBSERVATIONS			
		SUBFILE 1	SUBFILE 2		SUBFILES
		$O_1\ O_2 \ldots O_{n_1}$	$O_{n_1+1} \ldots O_{n_1+n_2}$	...	$O_{n_1+\ldots n_{s-1}+1} \ldots O_{n_1+\ldots+n_s}$
	V_1				
	V_2				
	.				
VARIABLES	.				
	.				
	V_p				

Scratch Data Sets

There are two data files which are used by the statistical data base. They are "DATA" and "BACKUP". DATA is the file which contains the most current form of your data matrix. It is updated upon completion of any procedure which modifies the data matrix or any variable names. Thus, DATA contains the data that will be used for any statistical calculations. BACKUP on the other hand, is not updated automatically. After the data has been first entered a copy of the DATA file is automatically put into BACKUP. From then on BACKUP can only be modified manually via the BACKUP PROCEDURE. This procedure will also let you retrieve the BACKUP file and copy it to the DATA file. So, if you erroneously alter your data matrix, the original data set is still retrievable.

Data File Configuration
The scratch file on the program medium, "DATA", and any files created to hold stored data and related information are configured as follows.

The data file is broken into logical records of 1280 bytes each (if you are unfamiliar with logical records, refer to your desktop's Programming Techniques Manual.) The first logical record is a "header file", which contains information pertinent to the data set which is stored in the remaining logical records. The header file contains the following information (variables):

	Limitations
data set title (T$)	80 characters
number of observations (No)	No*Nv <= 1500
number of variables (Nv)	50
variable names (Vn$(*))	10 characters each
number of subfiles(Ns)	20
subfile names(Sn$(*))	10 characters each
subfile characterizations (Sc(*))	N/A

The remaining logical records contain D(*,*), the data matrix.

For a detailed explanation of the data file, see the appendix.

Parser
BSDM is equipped with an elementary parser. This means that wherever an answer could require multiple responses the parser will separate your response into its individual parts. For example, when asked "What variables are desired?", you may respond in three ways:

1. ALL: enter ALL if you want the entire set of variables to be used
2. 1,2,3,...: enter the specific variables you want
3. 4-7: enter a dash (–) if you want all variables from 4 to 7

So, a sample response for the question might be:

1,3,5–8,10,15,21–25

The response would be interpreted to mean that you requested variables

1,3,5,6,7,8,10,15,21,22,23,24 and 25.

Thus, anywhere multiple values may be input, you may enter the responses in this manner.

In several cases the words "NONE" or "NO" are also possible responses. When they are allowed, it is mentioned in the prompt. These words may be used interchangeably.

Note

Entering negative numbers is no different than entering positive ones. For example, the input:

–10– –3,1–4

would mean all numbers between –10 and –3 and all between 1 and 4.

Incorrect Responses
If a response outside the range of plausible responses is input from the keyboard, an appropriate message is displayed on the CRT. Program execution is resumed by asking the question, or in some cases a previous question, again.

If a plausible response is given, but it is not correct, a couple of possibilities exist. First, if an incorrect value has been entered for a data point, it may be corrected using the EDIT program. Second, in many cases, responses to several questions are printed on the CRT. Then a question such as "Is the above information correct?" is asked. This allows any of the printed information to be changed.

Hardware Requirements

9826 or 9836 computer with 240k bytes, available user memory — required.

External printer — required. The CRT may be used as the printer but results will be difficult to read and understand.

External plotter — optional.

External mass storage — optional.

Note

Both the user-defined transformation option and non-linear regression require that you specify the form of the functions before you begin BSDM. See page 69 for an explanation.

Getting Started

1. If your 9826 or 9836 computer is ROM-based, go to Step 2. Otherwise, if your system is RAM-based, or if you do not wish to turn the computer **OFF** and the complete system is ready:

   a. Make sure that Basic is ready and all peripherals are properly connected and turned on. (Make sure P1 and P2 are set properly if a hardcopy plotter is being used).
   b. Insert the Basic Statistics disc into the internal flexible disc drive.
   c. Type: `Scratch A` (EXECUTE)
   d. Type: `Load "AUTOST",1` (EXECUTE)
   e. Go to Step 5.

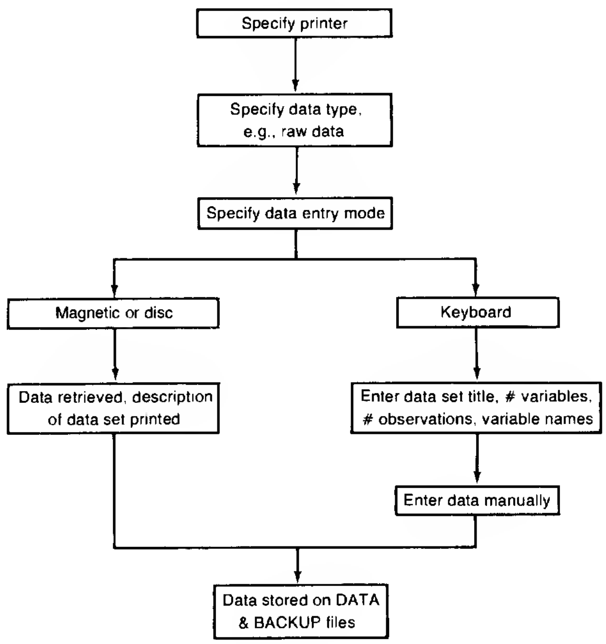
2. If the 9872C (or any peripheral) is being used, make sure it is properly connected and turned on. Make sure P1 and P2 are set properly if a hardcopy plotter is being used.
3. Insert the Basic Statistics disc into the internal disc drive.
4. Turn the computer on.
5. You will be asked a series of questions which should be self-explanatory. If you have any questions turn to the Special Considerations section of the manual covering the procedure in question. You will find some general comments on how that section of the program works.

Start

Object of Program

This program allows you to enter a data matrix into memory. The data may be entered from the keyboard, or from some other input device such as a graphics tablet, etc. Conversely, the data may have been entered previously and stored in the program scratch file ("DATA") or in a user-created file on a flexible disc or hard disc. In this case, the function of this program is to retrieve the previously stored data and place it into memory so that further operations can be performed. After the data is in memory, a listing option is available to obtain a complete or partial copy of the data.

Typical Program Flow



Special Considerations

Terminology

The displayed prompts concerning the scratch file ("DATA"), whether the data was stored by this program, and whether the data is in the proper configuration are explained here and in the Special Considerations section of General Information for BSDM.

The prompts concerning the data medium and program medium may cause confusion. The word "medium" is used since the set of programs making up this software package may be on floppy disc. Thus, the "program medium" refers to the disc on which the programs making up this package are stored. Conversely, the "data medium" refers to the disc on which the file containing the data matrix resides. In some cases, the program medium and the data medium are the same. However, this is not determined by the program and hence, the prompts are displayed to make sure the correct medium is in the correct device.

Data on Mass Storage

If the data is on a mass storage device, it may have been stored in one of four ways. The following discussion explains the prompts that apply to each situation.

1. If the data was entered using this statistics package (and was the last data set used on this package), it will be on the disc in the scratch file called "DATA". Thus, an affirmative answer to the prompt "Is data stored on the program medium's scratch file (DATA)?" will retrieve the data and related information.
2. The data may have been entered using the Basic Statistics and Data Manipulation routines and then stored using the STORE routine of BSDM. After specifying the file name and the storage unit in which the data resides, you should answer Yes to the prompt "Was data stored by this program?". Then, the data and related information will be retrieved.
3. The data may be stored as: all observations of variable one followed by all observations of variable two, etc. This is in the same configuration as data stored by the BSDM routines, i.e., variables = rows and observations = columns. To retrieve the data, a Yes response to the prompt "Is the data in proper configuration...?" should be given.
4. The data may be stored as: all variables of observation one followed by all variables of observation two, etc. This is the transpose of what is expected by the BSDM routines, i.e., observations = rows, variables = columns. To retrieve this type of data a Yes response should be given to the prompt "Data stored as contiguous array with observations = rows...?".

Notice that in cases 3 and 4, the data was stored by a program other than a statistics routine. Thus, no variable names or other auxiliary information will be stored along with the data.

As an example, suppose you have run your own program where you have created a file by storing data acquired from three sensors as it came in from the devices. A picture of five readings (observations) from the sensors would look like this:

Reading

	1	2	3	4	5
Sensor 1	7.2	7.4	7.1	7.2	7.3
Sensor 2	8.0	7.9	8.1	7.8	8.0
Sensor 3	7.8	7.5	7.5	7.6	7.9

If the data were stored in this order: 7.2, 7.4, 7.1, 7.2, 7.3, 8.0,..., 7.5, 7.6, 7.9, then it is in what we call the proper configuration, and the situation is that described in note 3 above.

Conversely, if the data were stored as: 7.2, 8.0, 7.8, 7.4, 7.9, 7.5, ... , 7.3, 8.0, 7.9, then it is the transpose of what is expected and the situation is that described in note 4 above.

Keyboard Entry
When entering data from the keyboard, an option to enter data one case at a time is offered. The following example will serve to explain this feature. Suppose an investigator has collected four observations on each of three variables. He has the following data matrix:

Variable

		1	2	3
	1	10	2	5
Observation	2	11	2	6
	3	9	3	7
	4	9	2	6

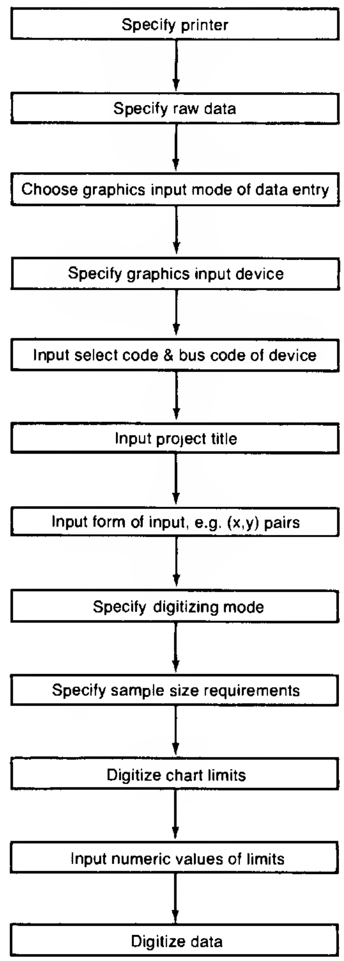
He elects to enter the data one case at a time. Then, when the prompt "Observation #, all variables (separated by commas) = ?" is displayed, he enters 10, 2, 5 and presses CONT, etc. This allows for quick entry of the data.

The other form of keyboard entry will prompt you at each observation for the required variable.

Missing Values
If you have missing values, use an unused number for a temporary code for a missing value. Subsequently you can change your values to the program's value of −9999999.99999 by using the TRANSFORM operation.

Graphics Input

Data may be input by digitizing from a graphics tablet. You may find this form of input very useful. The following diagram briefly describes the types of information requested by the program.



"Other" Input

Because of the wide variety of formats that could be used when entering data from "other" devices, no attempt was made to program in the necessary statements. It will be necessary for you to provide the statements before using the program. Refer to the Operating Manual of the appropriate device for detailed instructions. In general, though,

1. Type: `LOAD"FILE1"`
2. Press: (EXECUTE)
3. Type: `EDIT Other_input` (EXECUTE)
4. Change the 0 to a 1 in line `1731: Other_input: Implemented=0`
5. Press: (ENTER)
6. Press: (PAUSE)
7. Type: `EDIT Otherin` (EXECUTE)
8. Type in and enter the appropriate statements for "other" input, referring to the Operating Manual for the input device.

Edit

Object of Program

This program is designed to allow you to perform a variety of editing procedures on your data set. The editing capabilities include:

Correct a data value
Correct an entire observation
Delete a variable
Delete an observation
Add a variable
Add an observation
Insert an observation (in ordered data)
Delete a subfile

All of these operations may be performed repeatedly. For example, three variables may be added in succession. After the data matrix has been edited, you are given the option of listing the data.

Special Considerations

Order of Corrections

As stated in the program note printed on the screen, the data is renumbered after deletions or insertions are performed. For this reason, if more than one deletion (insertion) is to be performed, it is recommended that the highest-numbered observation (or variable) be deleted, then the next highest-numbered, etc. For example, if observations three and eight are to be deleted, then it is recommended to delete observation eight first, then observation three. Notice that if observation three were deleted, first, the subsequent renumbering would move observation eight to position seven. The recommendation is meant to alleviate confusion which may occur due to the renumbering. If you delete several observations at once using the answering technique described in the Special Considerations section of BSDM General Informations under "Parser", you do not need to worry about the renumbering problem. Your responses will be sorted from highest to lowest automatically. So to delete observations five through eight, just enter 5–8 and you will have no problems.

Subfiles
Insertions or deletions of observations will affect the content of subfiles which exist at the time of editing. For example, if subfile one consists of the first 10 observations while subfile two consists of the last 20 and if observation five is deleted, then observation ten (formerly numbered 11) will have jumped from subfile two to subfile one. Thus, it may be necessary to change the subfile structure after editing. It is recommended that subfiles be created only after all editing has been performed.

Correcting Data Value(s)
When correcting a data value, you must specify the variable number and observation number of the value to be corrected. Then, the old value is displayed prior to your correction so you can be sure you are altering the correct value.

Correcting Observation(s)
When correcting an entire observation, you specify the observation to be corrected. The old values are then listed on the screen and you may then enter the new values one-at-a-time.

Adding Observation(s)
In adding observations you will be asked to enter the number of observations that are to be placed at the end of the data matrix. Observations should be entered one-at-a-time with the data values separated by commas.

If an observation is to be inserted, the position of the insertion must be specified by entering the number of the existing observation which the insertion will precede. For example, if an observation is to be inserted between observations 8 and 9, you must enter 9 when the prompt "Insertion to precede observation #?" is displayed. You will then be asked to enter the number of observations that are to be inserted at this point.

Deleting Observation(s)
You will be asked to enter the numbers corresponding to the observations to be deleted. They will be sorted and the observations will be deleted from highest-numbered to lowest-numbered to avoid renumbering confusion.

Deleting Subfile(s)
This option works the same as deleting observations. All you need to specify is the subfile number and all observations within the subfile will be deleted. All observations after the ones deleted will be renumbered.

Deleting Variable(s)
You will be asked to enter the numbers corresponding to variables to be deleted. They will be sorted and the variables will be deleted from highest-numbered to lowest-numbered.

Exceeding Program Limitations
If the addition of an observation or of a variable will exceed program limitations, these options will not be executed.

Methods and Formulae
The data matrix is redimensioned into a row vector to facilitate the shuffling of elements necessitated by the editing operations. The vector contains all the observations of variable one, followed by the observations of variable two, etc. When an observation is inserted, for example, the elements of the data vector are shuffled one-at-a-time to make room for the incoming observation. Similarly, when an observation is deleted, the remaining observations are "packed" together so that the resultant data vector has no "holes" between observations.

Transform

Object of Program

This procedure is designed to allow you to transform your data. The transformations available fall into three categories. Algebraic transformations allow you to perform the standard algebraic operations on one or two variables in the data set. There is also the capability for you to define your own transformation. The second category of transformations is the assigning of missing values. With this section you may assign any value in the data set to correspond to missing data. The final section is new variables. Here, you may perform such operations as generating uniform random numbers, standardizing variables, lagging variables, creating rank variables, sequence variables, and variables corresponding to subfiles.

In all the sections the transformed results will be placed in a variable you specify, either old or newly-created. Hence, transformations on more than two variables may be performed iteratively or via a transformation defined by you.

Special Considerations

Missing Values (Algebraic Transformations)

None of the pre-specified algebraic transformations are applied to missing values. Thus, missing values are unaffected by these transformations. However, this is not necessarily the case with the user-defined transformation. If you define a transformation and there are missing values, you must make provisions to ensure that the transformation is not applied to the missing values (unless, of course, this is desired). This may be accomplished as explained below.

User-Defined (Algebraic Transformations)

Before you start to run the Basic Statistics and Data Manipulation program, you should prepare your own transformation function and store it on the data storage medium. Consider the following example. Suppose your data set consists of four variables. There are missing values. You desire to form variable five as the sum of the exponential of variables one and three. If there is a missing value in either of these variables, you wish to assign a missing value to the transformed variable. Recall that the data is of the form D(J,I) where J is the variable number and I is the observation number. In the transformation routine the variable Z is used to denote the variable where the transformed data is to be stored. Thus, to accomplish the above-described transformation, follow the instructions below:

1. Insert a flexible disc into the internal disc drive.
2. Type: `SCRATCH A` (EXECUTE)
3. Press: `EDIT` (EXECUTE)
4. Now you should be able to see line number = 10" on the upper-left corner of the CRT. Start to type in your function as a subroutine. Press (ENTER) after each line. For example:

```
10! A comment to identify perhaps your file name.
20  SUB Function (D(*),Z,I)
```

(Note: This line must be exactly the same as above.)

```
30  IF D(1,I)<>-9999999.99999 AND
D(3,I)<>-9999999.99999 THEN GO
40  D(Z,I)=-9999999.99999
```

```
50   GO TO 80
60   D(Z,I)=EXP(D(1,I))+EXP(E(3,I))
70 ! Note: The value of Z will be asked by the program. You must specify the
     variable numbers for the right hand side of the equation (i.e., 1 and 3)
80   SUBEND
```
(Note: This line must be the last line of the subroutine)

5. Press: CLR SCR
6. Type: `STORE` "your filename: mass storage identifier" EXECUTE

Now you can proceed with data entry through BSDM.

Declaring Missing Values

This section allows you to assign missing values to any or all of the variables in the data set. It may be used successively so that you can assign different missing values to each variable or different sets of variables. The program asks you to enter the variables to which a missing value is to be assigned. You are then asked what numbers are to be considered missing values for that group of variables. Then, these variables are scanned and all missing values are transformed to −9999999.99999, which is the standard missing value code.

Create Rank Variables

This operation will take a variable, rank its values in ascending order, and place the resulting ranks in the variable specified by you.

As an example, consider the following variable which has four observations.

Variable 1
23
25
29
20

You could create a second variable which contains the ranks corresponding to the observations in the first variable. You would obtain the following:

Variable 1	Variable 2
23	2
25	3
29	4
20	1

Creating Variables by Subfile

This option may only be used when a subfile structure is present. If used, this option will assign the subfile number associated with each observation to the specified variable.

For a simple example, suppose you have a data set with one variable containing five observations. Subfile one consists of the first two observations, while subfile two has the last three observations. In this case, you could create a second variable whose observations correspond to the subfile numbers associated with the original variable. This variable would look like the following.

Variable 2
1
1
2
2
2

Creating Variables by Sequence Number

By selecting this option, you can place the observation numbers in a specified variable. For example, in a data set with five observations, you could create a second variable which would look like the following:

Variable 2
1
2
3
4
5

Creating Standardized Score Variables

In this option, a chosen variable is standardized by the following formula:

$$\text{New Variable} = \frac{\text{Specified Variable} - \text{Mean of Specified Variable}}{\text{Standard Deviation of Specified Variable}}$$

The new variable can be placed in any variable you specify. Notice that standardized variables have a mean of zero and a standard deviation of one.

Creating Lag Variables

The lag variable operation will take the value of a chosen variable n-lags before and use it as the current observation of the lagged variable being created. As an example, consider the following data set:

		Var.1	Var.2
	1	2	3
	2	1	4
Obs.#	3	4	6
	4	1	2
	5	2	4

We can create variable 3 by lagging variable two by one lag. We can also create variable four by lagging variable one by two lags. We would obtain the following:

		Var.1	Var.2	Var.3	Var.4
	1	2	3	MV	MV
	2	1	4	3	MV
Obs. #	3	4	6	4	2
	4	1	2	6	1
	5	2	4	2	4

Notice that missing values are placed in the first n observations of an n-lag variable since lagged values cannot be assigned.

Creating Uniform Random Number Variables

This option allows you to generate uniform random numbers between zero and one and have them placed in a variable of your choice.

As an example of the use of this option, you could select a random sample of the observations in your data set to be used in a subsequent analysis. To do this, you could first use the uniform random number option to assign a uniform random number to each observation. Then, you could use the select procedure (described later in this manual) to chose a portion of the data set based on the uniform random numbers. For example, if you selected observations that had a corresponding random number value between zero and one-half, you expect to have selected about one-half of your data set.

Recode

Object of Program

This program allows you to assign codes to various categories or classes of data. The categories are intervals along the real number line and 20 of these may be specified. The recoding is done on one variable at a time. The same coding scheme may be used iteratively on successive variables. A summary of the coding intervals, codes, and number of observations assigned to each code is printed as hard copy.

Special Considerations

Coding Schemes

Four coding schemes are available for the sole purpose of eliminating unnecessary entries from the keyboard. If the coding intervals are all of the same length and are contiguous, that is, together they form a connected interval, then the interval construction can be accomplished internally knowing only the interval length and lower limit for the first interval. Similarly, if the intervals are of equal length but noncontiguous, for example,

[10,20),[25,35),[35,45),[50,60)

then the lower limit of each interval needs to be specified but the upper limit may be computed internally. Hence, the coding schemes are meant only to minimize the amount of information which needs to be entered from the keyboard. Clearly, the coding intervals could all be constructed by requiring you to enter the lower and upper limits for each and every interval (which is necessary, and what is done if the intervals are unequal and noncontiguous).

Coding is carried out one observation at at time. If you wish to recode more than one variable you must use the procedure successively, once for each variable to be recoded. Listed below are the available recoding options.

1. Contiguous intervals of equal length
2. Contiguous intervals of unequal length
3. Non-contiguous intervals of equal length
4. Non-contiguous intervals of unequal length

Option 1 will recode a variable into equally spaced intervals that are side by side. The second option will recode based on intervals of unequal length that are side by side. Options 3 and 4 will recode into intervals that need not be side by side. For equally spaced intervals, use option 3 and for unequally spaced intervals use option 4.

Brackets
The brackets used to denote the coding intervals are meant to follow their usual mathematical interpretation, that is, the intervals are closed on the left and open on the right. Hence, if you want a value to fall into a certain interval, make sure it is strictly less than the upper limit for the interval.

Observations Which Do Not Fall in an Interval
If an observation does not fall into any of the coding intervals, a table will appear giving you three options on how to handle these values. You may either 1) leave them unrecoded, 2) assign them a special code, or 3) assign them the missing value code.

Sort

Object of Program

This program allows the data matrix, or individual subfiles of the data matrix, to be sorted according to the values of one variable. For example, suppose you have five observations of three variables, say height, weight and age and want to arrange the observations in ascending order according to age. This is accomplished by sorting the data matrix according to variable three. The data may be sorted in ascending or descending order.

If you want to perform a hierarchical sort, the sort procedure must be used successively. For example, suppose you wish to sort a data set on weight and within weight by age. To do this, you should first sort on age and then use the sort procedure again and sort on weight. The sort procedure also sorts either in ascending or descending order. A sort in ascending order will place the observations in order from lowest to highest based on the variable sorted. A descending-order sort will put the observations in order from highest to lowest.

Special Considerations

Subfile Structure Options

If subfiles are ignored, the entire data set will be sorted and, in the process, the composition of the subfiles is subject to change. The option of sorting certain subfiles may be used to sort a single subfile or a set of successive subfiles according to one variable. The option of sorting all subfiles may be used to sort each and every subfile. The options of sorting certain subfiles and sorting all subfiles treat each subfile as if it were a separate data set. Thus, the sort is done with respect to one subfile at a time.

What Happens

It is important to note that entire observations are moved when the sort is carried out. Thus, referring to the example given in the Object of Program section above, a person's height and weight remain with the person's age as shown below.

Original Data Set

			Variable	
		Height	Weight	Age
	1	72	170	21
	2	70	165	25
Observation	3	69	150	20
	4	70	165	25
	5	73	160	19

Data Set Sorted by Age

			Variable	
		Height	Weight	Age
	1	73	160	19
	2	69	150	20
Observation	3	72	170	21
	4	70	165	25
	5	70	165	25

Subfiles

Object of Program

This program allows you to specify subfiles or logical groupings of the observations. This may be accomplished by entering the number of observations in each subfile or by entering the observation number of the first observation in each subfile. A third option is to create subfiles for each level of a specified variable. Names for the subfiles are entered in all cases. A fourth option allows you to destroy the existing subfile structure.

Special Considerations

Use of Subfiles

Subfiles may be created in order to specify logical groupings of observations. A subfile structure allows you to consider each subfile as a separate data set or to lump all the subfiles together and analyze the overall data set. For example, suppose you want to determine the output generated each day by each of three shifts. You would like to analyze the data separately for each of the three shifts as well as for the work force as a whole. You could form three separate data sets and do the individual analyses, then later join the three sets together for the overall analysis. However, since the same variables were measured for each of the shifts, the situation is well handled by specifying a subfile for each shift. The subfile structure options make it possible to do the analysis by subfile as well as for the overall data set.

Change Names

Object of Program

This program allows you to rename the data set, to rename variables and/or to rename subfiles. These names are then stored, along with the data, on the program medium's scratch file ("DATA"). You may change a single variable or subfile name, or you may change a set of names.

Store Data

Object of Program

This program allows you to store the entire data matrix and related information in a file so that it may be retrieved at a later date for further analysis. Alternatively, a subset of the data matrix may be stored by specifying which variables and/or subfiles are to be saved.

Special Considerations

Use of Program

The store feature will be useful in two different situations. First, if an investigator has a data set which he may want to analyze further at a later date, he may store it and retrieve it later via the Basic Statistics and Data Manipulation Start routine. Secondly, if several people have access to the data input programs, it becomes mandatory that each be able to store his data set in a unique place. Note that if only one person uses the routine on one data set it is unnecessary to use the store feature since the data and related information are kept in "DATA" – the scratch file on the program medium.

Protecting Existing Data

The existence of a file is checked in the program in an attempt to avoid the accidental loss of existing data. Thus, when a file is specified to receive the data, an attempt is made to ensure that you are not accidentally storing the new data in a file which you did not know existed.

List

Object of Program

This program allows you to obtain a listing of the data matrix. The listing will appear on the device that has been specified for hard-copy in the Start routine or in the Output Unit routine. You can list all the data, or a specified subset of the data. You may also specify how you want the data listed, i.e., by observation, by variable, etc.

Join

Object of Program

This progam allows you to join or combine two data sets into a single unit. One data set must be in memory and the other data set must have been previously stored by the Basic Statistics and Data Manipulation program. Two options are available. First, observations may be added together (if both sets have the same number of variables). Second, variables may be added together (if both sets have the same number of observations). A check is made in the program to make sure the two sets can be joined. Also, summary information on both data sets is printed before the joining operation is performed. Thus, the joining can be aborted if the resultant set will not be as expected.

Special Considerations

Adding Observations

Suppose data on six variables was gathered in each of the 52 weeks in 1975, analyzed, and stored on an auxiliary data disc. Suppose the same variables were measured in 1976, analyzed, and stored. If you are interested in lumping the two sets of data together for an overall analysis, you may use the Add Observations option of the joining routine. One set of data must be retrieved via the Start routine. Then, after entering the Join routine, the second set may be retrieved and the joining carried out. Notice that the variables must be in the same order in the two data sets.

Adding Variables

Suppose you measured five variables on each of 50 subjects in an experiment. These were analyzed and stored on disc. Later, you realize that three more variables are of interest. You measure these variables on the subjects in the same order as before and analyze them. All eight variables measured on each subject could be combined into a single data set via the joining routine.

Subfiles

If variables are added, the subfile structure assigned to the resultant data set is the subfile structure of data set #1, that is, the data set that is in machine memory prior to the joining operation. If observations are added, the following procedures are employed: 1) If no subfiles exist in either data set, the resultant data set has no subfiles. 2) If data set #1 has no subfiles, but data set #2 does, then a subfile named "SET #1" is created which consists of data set #1 and the subfiles of data set #2 remain unchanged. 3) If data set #1 contains subfiles, but data set #2 does not, then a subfile named "SET #2" is created which consists of data set #2 and the subfiles of data set #1 remain unchanged. 4) If both data sets contain subfiles, all of the subfiles of data set #1 are retained and as many subfiles of data set #2 are retained as possible – the upper limit of total subfiles for the resultant set being determined by the program limitations (see Special Considerations of Basic Statistics and Data Manipulation).

Printer Is

Object of Program

This program allows you to specify the device on which the hard-copy output will be printed, or conversely, to specify that no hard-copy is desired, i.e., that output be directed to the CRT.

Special Considerations

The hard copy option can be changed in two ways:

1. Select "PRINTER" key when you are asked to "SELECT ANY KEY".
2. This option can only be used when the program is not expecting an answer. For example, when Notes are displayed on the CRT and you are asked to press CONTINUE when ready. The printer may be changed as follows:

For Non-HP-IB Printer:

1. Type: `Hc=` (the select code of the desired printer) EXECUTE
2. Type: `Hcbus=999` EXECUTE

For HP-IB Printer:

1. Type: `Hc=` (the select code of the desired printer) EXECUTE
2. Type: `Hcbus=` (the bus address of the HP-IB device) EXECUTE

Select and Scan

Object of Program

This program allows you to look at a portion of your data set that satisfies a conditional statement. If you are scanning the data set, your output will include the observation numbers satisfying the scanning criterion and their distribution throughout the subfile structure. The data set which you are scanning will remain unaltered. When using the select option, your output will be the same as scanning, but the data set will be reduced to just those observations satisfying the selection criterion. Remember, the BACKUP file (explained in Special Considerations of Basic Statistics and Data Manipulation) will contain the original data set. The selection and scanning procedure may be performed over all subfiles or over a user-specified subset of the data.

Special Considerations

There are four different scanning or selection criteria offered in this routine. Explanations of each conditional statement follow.

One Variable

This option will allow you to "edit" your data set based on specified values for one chosen variable. For example, you may scan (or select from) your data set based on variable number two and have the routine report the observations where variable two has any of the following values: 1, 2.6, 4–8.

Variable A OR Variable B

This option will allow you to "edit" your data set based on specified values of two chosen variables. An OR operation links the two variables. For example, if two of your variables are temperature and humidity, you may want to select (or scan) all observations that have a temperature of 70–80 degrees, OR have a humidity level of 50–80.

Variable A AND Variable B

This option performs much like the OR option except is uses an AND operator. For example, you may want to select (or scan) all observations that have a temperature of 72 degrees AND a humidity level of 50–80.

Variable A = Variable B
In this case the observations that would be selected (or scanned) are the observations where Variable A has the same value as Variable B. For example, you might want to know which observations have equal temperature and humidity level.

Basic Statistics

Object of Program

This program computes a variety of summary statistics for data which was entered via the Start routine of Basic Statistics and Data Manipulation. The statistics may be computed by subfile or for the entire data set (ignoring subfiles). Basic statistics which are computed include: number of observations, number of missing values, sum, mean, variance, standard deviation, coefficient of skewness, coefficient of kurtosis, coefficient of variation, standard error of the mean, and a confidence interval on the mean. An option is available to compute a correlation matrix for data sets having more than one variable. Order statistics computed include: the maximum, the minimum, range, and midrange. Additional order statistics which may be computed include:the median, 25th percentile, 75th percentile, Tukey's middlemeans, and user-specified percentiles. These statistics are divided into three groups. You may specify any or all of the groups for output.

Special Considerations

Parser on Statistics Options

Three options for statistics will be offered. They are 1) the common summary statistics, 2) the correlation matrix, and 3) the order statistics such maximum minimum, median, etc. You may respond "ALL" to the prompt asking you for your choice of options. Or, you may choose a portion of the options by responding as documented in the General Information section of Basic Statistics and Data Manipulation e.g., 1–2.

Data Type

If the data input type is not "RAW DATA", the Basic Statistics may not be computed. For example, Basic Statistics cannot be computed if the covariance matrix was entered as data.

Hard-Copy Output

If a hard copy of the statistics is not being made, the program halts occasionally so that you may study the results on the CRT. In this case, simply press CONTINUE to continue program execution.

Additional Order Statistics

If the option to obtain additional order statistics (Tukey's middlemeans and percentiles) is exercised, the data matrix is sorted and the observations of each variable are arranged in ascending order. At the end of the program the original data matrix is re-loaded into memory. Thus, if the program is aborted, that is, if the program is stopped before the reloading can occur, the data matrix will be in the sorted state. So, if the portion of the program used to calculate additional order statistics is accessed, abortion of the program is discouraged.

Methods and Formulae

Variance: The best unbiased estimator is calculated by these programs, i.e., the denominator in the formula is N–1, where N is the number of observations used in the calculation.

Correlations: Suppose you have the following data matrix:

		OBSERVATION 1	2	3	4	5
VARIABLE	1	5	M	3	4	5
	2	6	7	M	6	4
	3	1	3	2	1	1

Here, an M denotes a missing value. When computing the correlation between variables 1 and 2, we discard observations 2 and 3 since variable 1 is missing a data value for observation 2 and variable 2 is missing the data value for observation 3. However, when computing the correlation between variables 1 and 3, we need only discard observation 2. Similarly, the correlation between variables 2 and 3 is computed by discarding observation 3. Hence, the correlations may be based on different numbers of observations. An observation is thrown out if a data value from that observation is missing from one of the two variables for which the correlation is being computed.

Tukey's Middlemeans

Midmean: The midmean is the sum of all observations between (and including, if applicable) the 25th and 75th percentiles divided by the number of observations between those two percentiles. That is, it is the mean of all observations between the 25th and 75th percentiles.

Trimean: The trimean is a weighted average of the median and the 25th and 75th percentiles:

(1/4)(25th percentile + 2(median) + 75th percentile).

Midspread: The midspread is the difference between the 75th and 25th percentiles:

75th percentile – 25th percentile.

Go To Advanced Stat

Objective

This procedure loads a file which prompts you to remove the BSDM program medium and insert the desired advanced statistics program medium into the mass storage device. You press CONTINUE after you have made this change. The new routines are then prepared to carry on further analyses on the data set in memory.

Return To BSDM

Objective

This procedure operates in the reverse of "Go To Advanced Stat" and should be used when you wish to return to the BSDM routines from an advanced statistics routine.

Backup

Objective

This routine allows you to transfer the original data which is stored in the file called "BACK-UP" to the program scratch file called "DATA". You might find this useful in a case where the data currently in the "DATA" file is not the data you wish to be analyzing. This could occur, for example, if you inadvertantly stored a transformed variable in place of one of your original variables. Note that no operations, including editing, are performed on the data stored on the "BACKUP" file.

This routine also allows you to transfer the data set in the opposite direction. That is, you may transfer the data stored in "DATA" to the "BACKUP" file. You might choose to do this after you have edited the original data set but before you perform any other operations. Then, the "BACKUP" file would contain the corrected original data without any further manipulations or modifications.

Examples

Example 1

This is a hypothetical set of data from a non-existent factory. The purpose of this example is to show the use, in part, of the LIST, EDIT TRANSFORM, SORT, SUBFILE, and STATS routines.

```
BASIC STATISTICS AND DATA MANIPULATION


[Answer all yes/no questions with Y/N]


Are you going to user defined transformation or do Non-linear regression ?   (Y/N)

N
Are you using an HPIB Printer?

YES
Enter select code, bus address (if 7,1 press CONT) ?
```

We input these values separated by a comma or press CONTINUE if default (7,1) is correct

```
*************************************************************************************
*                                 DATA MANIPULATION                                 *
*************************************************************************************


Enter DATA TYPE:

1
```

Raw data

```
Mode number = ?

1
```

The data will be entered by typing it in on the keyboard.

```
Project title for this data set (<= 80 characters) = ?

HYPOTHETICAL FACTORY DATA
```

Title

```
Number of variables =

5
```

Nv

```
Number of observations/variable = ?

17
```

No

```
Variable # 1 name (<= 10 characters) =
?

TEMP(C)
```

Label for variable 1

```
Variable # 2 name (<= 10 characters) =
?

PRODUCTION
```

Label for variable 2

```
Variable # 3 name (<= 10 characters) =
?

DAYS
```

Label for variable 3

```
Variable # 4 name (<= 10 characters) =
?

PAYROLL
```

Label for variable 4

```
Variable # 5 name (<= 10 characters) =
?

WATER USE
```

Label for variable 5

```
Is above information correct?

YES
```

Approve information on CRT (shown below).

HYPOTHETICAL FACTORY DATA

```
Data file name:

Data type is:   Raw data

Number of observations:    17
Number of variables:        5

Variable names:
   1. TEMP(C)
   2. PRODUCTION
   3. DAYS
   4. PAYROLL
   5. WATER USE

Do you want to enter data one case at a time, i.e., by observation?
YES
Observation # 1 , all variables (separated by commas) =
?
14.9,6396,21,134,3373
Observation # 2 , all variables (separated by commas) =
?
18.4,5736,22,146,3110
Observation # 3 , all variables (separated by commas) =
?
21.6,6116,22,158,3180
Observation # 4 , all variables (separated by commas) =
?
25.2,8287,20,171,3293
Observation # 5 , all variables (separated by commas) =
?
26.3,13313,25,198,3390
Observation # 6 , all variables (separated by commas) =
?
27.2,13108,23,194,4287
Observation # 7 , all variables (separated by commas) =
?
22.2,10768,20,180,3852
Observation # 8 , all variables (separated by commas) =
?
17.1,12173,23,191,3366
Observation # 9 , all variables (separated by commas) =
?
12.5,11390,20,195,3532
Observation # 10 , all variables (separated by commas) =
?
6.9,12707,20,192,3614
Observation # 11 , all variables (separated by commas) =
?
6.4,15022,22,200,3896
Observation # 12 , all variables (separated by commas) =
?
13.3,13114,19,211,3437
Observation # 13 , all variables (separated by commas) =
?
18.2,12257,22,203,3324
Observation # 14 , all variables (separated by commas) =
?
22.8,13118,22,197,3214
Observation # 15 , all variables (separated by commas) =
?
26.1,13100,21,196,4345
Observation # 16 , all variables (separated by commas) =
```

All variables will be entered separately by commas.

```
?
26.3,16716,21,205,4936
Observation # 17 , all variables (separated by commas) =
?
4.2,14056,22,205,3624
PROGRAM NOW STORING DATA ON SCRATCH DATA FILE AND BACKUP FILE

SELECT ANY KEY
```

LIST

Select Special Function Key-LIST

```
Option number = ?

1
Enter method for listing data:
```

List all the data

```
3
```

In tabular form

```
HYPOTHETICAL FACTORY DATA

Data type is: Raw data
```

OBS#	Variable # 1 (TEMP(C))	Variable # 2 (PRODUCTION)	Variable # 3 (DAYS)	Variable # 4 (PAYROLL)	Variable # 5 (WATER USE)
1	14.90000	6396.00000	21.00000	134.00000	3373.00000
2	18.40000	5736.00000	22.00000	146.00000	3110.00000
3	21.60000	6116.00000	22.00000	158.00000	3180.00000
4	25.20000	8287.00000	20.00000	171.00000	3293.00000
5	26.30000	13313.00000	25.00000	198.00000	3390.00000
6	27.20000	13108.00000	23.00000	194.00000	4287.00000
7	22.20000	10768.00000	20.00000	180.00000	3852.00000
8	17.10000	12173.00000	23.00000	191.00000	3366.00000
9	12.50000	11390.00000	20.00000	195.00000	3532.00000
10	6.90000	12707.00000	20.00000	192.00000	3614.00000
11	6.40000	15022.00000	22.00000	200.00000	3896.00000
12	13.30000	13114.00000	19.00000	211.00000	3437.00000
13	18.20000	12257.00000	22.00000	203.00000	3324.00000
14	22.80000	13118.00000	22.00000	197.00000	3214.00000
15	26.10000	13100.00000	21.00000	196.00000	4345.00000
16	26.30000	16716.00000	21.00000	205.00000	4936.00000
17	4.20000	14056.00000	22.00000	205.00000	3624.00000

```
Option number = ?
0
SELECT ANY KEY
```

Exit List routine

EDIT ROUTINES

Select Special Function Key-EDIT

```
Select option desired :

1
Observation number (enter 'NONE' when done) = ?
```

Choose to correct a data value.

```
11
Variable number = ?
```

At observation #11

```
2
Old value =  15022 -- Correct value =
?
```

For variable 2

```
15024
```

Should be 15024

```
 OBS      VAR          OLD              NEW
  #        #          VALUE            VALUE
  11       2      15022.00000      15024.00000
Observation number (enter 'NONE' when done) =
```

```
NONE
Select option desired :

6                                                                     Delete an observation
Which observations are to be deleted ?
10
Observation # 10  deleted.
 16 observations remain.
Select option desired :

4                                                                     Add an observation
Are observations ordered, i.e., should additions be inserted?

NO                                                                    Add at the end
How many observations are to be added?

1
Enter observation # 17  (variables separated by commas) :
?
4.2,12707,20,192,3614
Observation # 17 Variable # 1  =  4.2
Observation # 17 Variable # 2  =  12707
Observation # 17 Variable # 3  =  20                                  New observation #17
Observation # 17 Variable # 4  =  192
Observation # 17 Variable # 5  =  3614

  Total number of observations now =  17
Select option desired :

0                                                                     Exit Edit routines
PROGRAM NOW UPDATING SCRATCH DATA FILE
SELECT ANY KEY
                                                LIST
                                                                      Select Special Function Key-LIST
Option number = ?

1                                                                     List all the data
Enter method for listing data:

3                                                                     In tabular form
```

HYPOTHETICAL FACTORY DATA

Data type is: Raw data

OBS#	Variable # 1 (TEMP(C))	Variable # 2 (PRODUCTION)	Variable # 3 (DAYS)	Variable # 4 (PAYROLL)	Variable # 5 (WATER USE)
1	14.90000	6396.00000	21.00000	134.00000	3373.00000
2	18.40000	5736.00000	22.00000	146.00000	3110.00000
3	21.60000	6116.00000	22.00000	158.00000	3180.00000
4	25.20000	8287.00000	20.00000	171.00000	3293.00000
5	26.30000	13313.00000	25.00000	198.00000	3390.00000
6	27.20000	13108.00000	23.00000	194.00000	4287.00000
7	22.20000	10768.00000	20.00000	180.00000	3852.00000
8	17.10000	12173.00000	23.00000	191.00000	3366.00000
9	12.50000	11390.00000	20.00000	195.00000	3532.00000
10	6.40000	15024.00000	22.00000	200.00000	3896.00000
11	13.30000	13114.00000	19.00000	211.00000	3437.00000
12	18.20000	12257.00000	22.00000	203.00000	3324.00000
13	22.80000	13118.00000	22.00000	197.00000	3214.00000
14	26.10000	13100.00000	21.00000	196.00000	4345.00000
15	26.30000	16716.00000	21.00000	205.00000	4936.00000
16	4.20000	14056.00000	22.00000	205.00000	3624.00000
17	4.20000	12707.00000	20.00000	192.00000	3614.00000

Option number = ?	
0	Exit List routine
SELECT ANY KEY	
	Select Special Function Key labeled-STORE
Enter option number desired :	
1	Store all the data
Name of data file = ?	
HYPO:INTERNAL	On this file on our floppy
Is data medium placed in device INTERNAL	
?	
YES	
PROGRAM NOW STORING DATA ON HYPO:INTERNAL	
Is program medium replaced in device?	
YES	
Enter option number desired :	
0	Exit Store routine
SELECT ANY KEY	

TRANSFORMATION ROUTINES

Select option desired :	Select Special Function Key labeled-TRANSFORM
1	Algebraic transformations
Transformation number = ?	
1	$a*(X \uparrow b)+c$
Variable number corresponding to X = ?	
5	
Parameter a = ?	
.2642	To convert liters to gallons
Parameter b = ?	$X_6 = .2642X_5$
1	
Parameter c = ?	
0	
Store transformed data in Variable # (<= 6)	
?	
6	
Variable name (<= 10 characters) = ?	
GALLONS	X_6 now called GALLONS.
Is above information correct?	
YES	
press 'CONTINUE' when ready	

```
The following transformation was performed: a*(X^b)+c
  where a =  .2642
        b =  1
        c =  0
        X is Variable # 5
        Transformed data is stored in Variable # 6 (GALLONS).
```

Select option desired :	
0	Exit transformation routine
PROGRAM NOW UPDATING SCRATCH DATA FILE	
SELECT ANY KEY	

SORT ROUTINES

```
ENTER OPTION NUMBER DESIRED :                              Select Special Function Key labeled-SORT

1                                                          Sort in ascending order
Number of the Variable on which to sort =

3                                                          On variable 3 (Days in month)
Data set:
                        HYPOTHETICAL FACTORY DATA
has been arranged in ascending order according to Variable # 3

ENTER OPTION NUMBER DESIRED :

0                                                          Exit sort routine
PROGRAM NOW UPDATING SCRATCH DATA FILE
SELECT ANY KEY                    LIST
                                                           Select Special Function Key labeled-LIST
Option number = ?

1                                                          List all the data
Enter method for listing data:

3                                                          In tabular form
```

HYPOTHETICAL FACTORY DATA

Data type is: Raw data

OBS#	Variable # 1 (TEMP(C))	Variable # 2 (PRODUCTION)	Variable # 3 (DAYS)	Variable # 4 (PAYROLL)	Variable # 5 (WATER USE)
1	13.30000	13114.00000	19.00000	211.00000	3437.00000
2	22.20000	10768.00000	20.00000	180.00000	3852.00000
3	25.20000	8287.00000	20.00000	171.00000	3293.00000
4	4.20000	12707.00000	20.00000	192.00000	3614.00000
5	12.50000	11390.00000	20.00000	195.00000	3532.00000
6	26.30000	16716.00000	21.00000	205.00000	4936.00000
7	26.10000	13100.00000	21.00000	196.00000	4345.00000
8	14.90000	6396.00000	21.00000	134.00000	3373.00000
9	6.40000	15024.00000	22.00000	200.00000	3896.00000
10	21.60000	6116.00000	22.00000	158.00000	3180.00000
11	18.20000	12257.00000	22.00000	203.00000	3324.00000
12	22.80000	13118.00000	22.00000	197.00000	3214.00000
13	18.40000	5736.00000	22.00000	146.00000	3110.00000
14	4.20000	14056.00000	22.00000	205.00000	3624.00000
15	27.20000	13108.00000	23.00000	194.00000	4287.00000
16	17.10000	12173.00000	23.00000	191.00000	3366.00000
17	26.30000	13313.00000	25.00000	198.00000	3390.00000

OBS#	Variable # 6 (GALLONS)
1	908.05540
2	1017.69840
3	870.01060
4	954.81880
5	933.15440
6	1304.09120

```
   7     1147.94900
   8      891.14660
   9     1029.32320
  10      840.15600
  11      878.20080
  12      849.13880
  13      821.66200
  14      957.46080
  15     1132.62540
  16      889.29720
  17      895.63800

Option number = ?
0
```

Exit list routine

```
SELECT ANY KEY
```

SUBFILE

Select Special Function Key labeled-SUBFILES

```
Option number = ?

2
Number of subfiles ( <=20 ) = ?
```

Select method of subfile specifications which ask you to enter the first observation in each subfile.

```
2
Name of Subfile # 1 ( <= 10 characters ) =
?

FY'76
Name of Subfile # 2 ( <= 10 characters ) =
?

FY'77
Subfile # 2 : number of first observation =
?

13
Is the above information correct ?
YES

Subfile name:     beginning observation--number of observations
 1. FY'76                       1                        12
 2. FY'77                      13                         5
```

Summary

```
Option number = ?
0
```

Exit subfile routine

```
PROGRAM NOW STORING DATA
```

BASIC STATISTICS ROUTINES

```
SELECT ANY KEY
```

Select Special Function Key labeled-STATS

```
What statistic options are desired ?

1
```

Mean, Ci, Variance, Standard Deviation, Skewness, Kurtosis

```
VARIABLES
?
ALL
```

Compute statistics for all variables

```
Confidence coefficient for confidence interval on the mean(e.g. 90,95,99%) = ?
95
Option number = ?

2
```

Compute statistics for selected subfiles.

```
What subfiles are desired ?
1
```

For FY'76

```
**************************************************************************
*                          SUMMARY STATISTICS                            *
*                            ON DATA SET:                                *
*                      HYPOTHETICAL FACTORY DATA                         *
**************************************************************************
```

Subfile: FY'76

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
TEMP(C)	12	0	213.7000	17.8083	56.9572	7.5470
PRODUCTION	12	0	138993.0000	11582.7500	10478676.7500	3237.0784
DAYS	12	0	250.0000	20.8333	1.0606	1.0299
PAYROLL	12	0	2242.0000	186.8333	504.5152	22.4614
WATER USE	12	0	43996.0000	3666.3333	274270.7879	523.7087
GALLONS	12	0	11623.7432	968.6453	19144.5508	138.3638

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
TEMP(C)	42.37903	2.17863	13.01195	22.60471
PRODUCTION	27.94741	934.46405	9525.47409	13640.02591
DAYS	4.94332	.29729	20.17882	21.48784
PAYROLL	12.02217	6.48405	172.55832	201.10834
WATER USE	14.28426	151.18168	3333.49825	3999.16841
GALLONS	14.28426	39.94220	880.71024	1056.58030

VARIABLE	SKEWNESS	KURTOSIS
TEMP(C)	-.53473	-.96332
PRODUCTION	-.42217	-.66250
DAYS	-.18352	-1.18041
PAYROLL	-1.22848	.55306
WATER USE	1.34739	.89749
GALLONS	1.34739	.89749

```
What statistic options are desired ?                    Mean, Ci, Variance, Standard Deviation,
1                                                       Skewness, Kurtosis
VARIABLES=?
ALL                                                     Compute statistics for all variables
Confidence coefficient for confidence interval on the mean(e.g. 90,95,99) = ?
95
Option number = ?
2                                                       Compute statistics for selected subfiles.
What subfiles are desired ?
2                                                       For subfile FY'77
```

```
*******************************************************************************
*                            SUMMARY STATISTICS                               *
*                              ON DATA SET:                                   *
*                        HYPOTHETICAL FACTORY DATA                            *
*******************************************************************************
-------------------------------------------------------------------------------
Subfile: FY'77
-------------------------------------------------------------------------------
```

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
TEMP(C)	5	0	93.2000	18.6400	85.7230	9.2587
PRODUCTION	5	0	58386.0000	11677.2000	11481348.7000	3388.4139
DAYS	5	0	115.0000	23.0000	1.5000	1.2247
PAYROLL	5	0	934.0000	186.8000	547.7000	23.4030
WATER USE	5	0	17777.0000	3555.4000	200388.8000	447.6481
GALLONS	5	0	4696.6834	939.3367	13987.4669	118.2686

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
TEMP(C)	49.67099	4.14060	7.14334	30.13666
PRODUCTION	29.01735	1515.34476	7469.74622	15884.65378
DAYS	5.32498	.54772	21.47921	24.52079
PAYROLL	12.52837	10.46614	157.74009	215.85991
WATER USE	12.59065	200.19431	2999.54742	4111.25258
GALLONS	12.59065	52.89134	792.48043	1086.19293

VARIABLE	SKEWNESS	KURTOSIS
TEMP(C)	-.68247	-.77608
PRODUCTION	-1.35662	.05662
DAYS	.91287	-.50000
PAYROLL	-1.30917	.02054
WATER USE	.91055	-.44827
GALLONS	.91055	-.44827

```
What statistic options are desired ?

2                                        Correlation matrix
VARIABLES=
?
ALL                                      Compute statistics for all variables
Option number = ?

2                                        Compute statistics on selected subfiles.
What subfiles are desired ?
1,2
```

```
****************************************************************************
*                           SUMMARY STATISTICS                             *
*                              ON DATA SET:                                *
*                        HYPOTHETICAL FACTORY DATA                         *
****************************************************************************
----------------------------------------------------------------------------
Subfile: FY'76
----------------------------------------------------------------------------
                            CORRELATION MATRIX


             PRODUCTION      DAYS     PAYROLL  WATER USE    GALLONS
TEMP(C)       -.1113482  .1627763  -.1009200   .2511888   .2511888
PRODUCTION               .0081945   .8872541   .6589095   .6589095
DAYS                               -.1113502  -.0368011  -.0368011
PAYROLL                                        .3820119   .3820119
WATER USE                                                1.0000000


----------------------------------------------------------------------------
----------------------------------------------------------------------------
Subfile: FY'77
----------------------------------------------------------------------------
                            CORRELATION MATRIX


             PRODUCTION      DAYS     PAYROLL  WATER USE    GALLONS
TEMP(C)       -.0709995  .6614042  -.1292917   .2656162   .2656162
PRODUCTION               .4116924   .9974909   .5754985   .5754985
DAYS                                .3924963   .0209757   .0209757
PAYROLL                                        .5259584   .5259584
WATER USE                                                1.0000000


----------------------------------------------------------------------------
What statistic options are desired ?
3
VARIABLES
?
ALL
Option number = ?
2
What subfiles are desired ?
1,2
```

Median, Mode, Percentiles, Min, Max, Range. (3)

Compute statistics for all variables (ALL)

Compute statistics for selected subfiles. (2)

Both subfiles (1,2)

```
***************************************************************************
*                           SUMMARY STATISTICS                            *
*                              ON DATA SET:                               *
*                        HYPOTHETICAL FACTORY DATA                        *
***************************************************************************
```

Subfile: FY'76

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
TEMP(C)	26.30000	4.20000	22.10000	15.25000
PRODUCTION	16716.00000	6116.00000	10600.00000	11416.00000
DAYS	22.00000	19.00000	3.00000	20.50000
PAYROLL	211.00000	134.00000	77.00000	172.50000
WATER USE	4936.00000	3180.00000	1756.00000	4058.00000
GALLONS	1304.09120	840.15600	463.93520	1072.12360

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
TEMP(C)	19.90000	12.90000	24.00000
PRODUCTION	12482.00000	9527.50000	13116.00000
DAYS	21.00000	20.00000	22.00000
PAYROLL	195.50000	175.50000	201.50000
WATER USE	3484.50000	3308.50000	3874.00000
GALLONS	920.60490	874.10570	1023.51080

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
TEMP(C)	18.83333	19.17500	11.10000
PRODUCTION	12222.66667	11901.87500	3588.50000
DAYS	20.83333	21.00000	2.00000
PAYROLL	193.33333	192.00000	26.00000
WATER USE	3522.00000	3537.87500	565.50000
GALLONS	930.51240	934.70658	149.40510

Other percentiles(Y/N)?
NO

Subfile: FY'77

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
TEMP(C)	27.20000	4.20000	23.00000	15.70000
PRODUCTION	14056.00000	5736.00000	8320.00000	9896.00000
DAYS	25.00000	22.00000	3.00000	23.50000
PAYROLL	205.00000	146.00000	59.00000	175.50000
WATER USE	4287.00000	3110.00000	1177.00000	3698.50000
GALLONS	1132.62540	821.66200	310.96340	977.14370

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
TEMP(C)	18.40000	17.10000	18.40000
PRODUCTION	13108.00000	12173.00000	13108.00000
DAYS	23.00000	22.00000	23.00000
PAYROLL	194.00000	191.00000	194.00000
WATER USE	3390.00000	3366.00000	3390.00000
GALLONS	895.63800	889.29720	895.63800

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
TEMP(C)	20.60000	18.07500	1.30000
PRODUCTION	12864.66667	12874.25000	935.00000
DAYS	22.66667	22.75000	1.00000
PAYROLL	194.33333	193.25000	3.00000
WATER USE	3460.00000	3384.00000	24.00000
GALLONS	914.13200	894.05280	6.34080

Other percentiles?
NO

What statistic options are desired ?
0 — Exit basic statistics routine
SELECT ANY KEY

Note: All Basic Statistics for these subfiles could have been obtained more efficiently than we demonstrated in this example by responding "ALL" to the above question.

Example 2

The data set is from the MINITAB STUDENT HANDBOOK authored by T. Ryan, and B. Joiner and published by the Duxbury Press (1976). The data appeared on page 279. The operation performed on two sets SAMPLE A and SAMPLE B demonstrate the following operations: JOIN, LIST, RECODE, SUBFILE (by variable), STORE, SELECT, and STATS.

```
BASIC STATISTICS AND DATA MANIPULATION

[Answer all yes/no questions with Y/N]

Are you going to use user defined transformation or non-linear regression ?   (Y/N)
NO
Are you using an HPIB Printer?

YES
Enter select code, bus address (if 7,1 press CONT) ?


*****************************************************************************************
*                                  DATA MANIPULATION                                    *
*****************************************************************************************


Enter DATA TYPE:

1
Mode number = ?

2
Is data stored on the program's scratch file (DATA)?

NO
Data file name = ?

GRADEB:INTERNAL
Was data stored by the BS&DM system ?

YES
Is data medium placed in device INTERNAL
?

YES
Is program medium placed in correct device ?

YES
PROGRAM NOW STORING DATA ON SCRATCH DATA FILE AND BACKUP FILE



                                       SAMPLE B


Data file name: GRADEB:INTERNAL

Data type is:    Raw data

Number of observations:     50
Number of variables:         3
```

Annotations:

- (beside "1 / Mode number"): Raw data
- (beside "2"): Data is from mass storage
- (beside "GRADEB:INTERNAL"): The data was stored under the name GRADEB in a different place, so the program must retrieve it.
- (beside "Data file name: GRADEB:INTERNAL"): This data is the second set of 50 student grades (GPA) and scores on the ACT tests (Verb and Math). The data taken from the Minitab Student Handbook on page 279.

```
Variable names:
   1. VERB
   2. MATH
   3. GPA

Subfiles:  NONE

SELECT ANY KEY

Option number = ?
```

Select Special Function Key labeled-LIST

```
1
Enter method for listing data:
```

List all the data.

```
3
```

In tabular form.

SAMPLE B

Data type is: Raw data

OBS#	Variable # 1 (VERB)	Variable # 2 (MATH)	Variable # 3 (GPA)
1	500.00000	661.00000	2.30000
2	460.00000	692.00000	1.40000
3	717.00000	672.00000	2.80000
4	592.00000	441.00000	2.40000
5	752.00000	729.00000	3.40000
6	695.00000	681.00000	2.50000
7	610.00000	777.00000	3.60000
8	620.00000	638.00000	2.60000
9	682.00000	701.00000	3.60000
10	524.00000	700.00000	2.90000
11	552.00000	692.00000	2.60000
12	703.00000	710.00000	3.80000
13	584.00000	738.00000	3.00000
14	550.00000	638.00000	2.50000
15	659.00000	672.00000	3.50000
16	585.00000	605.00000	2.00000
17	578.00000	614.00000	3.00000
18	533.00000	630.00000	2.00000
19	532.00000	586.00000	1.80000
20	708.00000	701.00000	2.30000
21	537.00000	681.00000	2.10000
22	635.00000	647.00000	3.00000
23	591.00000	614.00000	3.30000
24	552.00000	669.00000	3.00000
25	557.00000	674.00000	3.20000
26	599.00000	664.00000	2.30000
27	540.00000	658.00000	3.30000
28	752.00000	737.00000	3.30000
29	726.00000	800.00000	3.90000
30	630.00000	668.00000	2.10000
31	558.00000	567.00000	2.60000
32	646.00000	771.00000	2.40000
33	643.00000	719.00000	3.30000
34	606.00000	755.00000	3.10000
35	682.00000	652.00000	3.60000
36	565.00000	672.00000	2.90000
37	578.00000	629.00000	2.40000
38	488.00000	611.00000	1.80000
39	361.00000	602.00000	2.40000
40	560.00000	639.00000	2.90000
41	630.00000	647.00000	3.50000

42	666.00000	705.00000	3.40000
43	719.00000	668.00000	2.30000
44	669.00000	701.00000	2.90000
45	571.00000	647.00000	1.80000
46	520.00000	583.00000	2.80000
47	571.00000	593.00000	2.30000
48	539.00000	601.00000	2.50000
49	580.00000	630.00000	2.40000
50	629.00000	695.00000	2.90000

```
Option number = ?

0
SELECT ANY KEY
```

Exit List routine.

JOIN ROUTINE

Select Special Function Key labeled-JOIN

```
Option number = ?
2
```

Choose to add observations.

```
Do you wish to continue with the JOIN procedure ?

Title for combined data set (<= 80 characters) = ?
```

To continue you must have

1. Data Set #1 currently in memory.
2. Data Set #2 previously stored by this program.
3. Total observations times varibles < 1500.
4. Each data set must contain the same number of variables arranged in the same order.

```
TOTAL ACT SCORE/GPA COMPARISON DATA
File name of data set #2 = ?

GRADEA:INTERNAL
Is data set #2 medium placed in device INTERNAL
?

YES
Press 'CONTINUE' when ready to continue.
Press 'CONTINUE' when ready to continue.
Is program medium placed in device ?

YES
```

This data set (the first set A in the Minitab manual) was previously stored.

```
                    TOTAL ACT SCORE/GPA COMPARISON DATA

Number of variables: 3
Number of observations: 100

Variable names:
 1. VERB
 2. MATH
 3. GPA
Subfiles: NONE
PROGRAM NOW UPDATING SCRATCH DATA FILE
Option number = ?
```

The two data sets are combined. That is the second 50 observations are 'attached' to the bottom of the original 50 observations.

```
0
SELECT ANY KEY
```

Exit Join routine

LIST ROUTINE

Select Special Function Key labeled-LIST

```
Option number = ?

1
Enter method for listing data:
```

List all the data

```
3
```

In tabular form

TOTAL ACT SCORE/GPA COMPARISON DATA

Data type is: Raw data

OBS#	Variable # 1 (VERB)	Variable # 2 (MATH)	Variable # 3 (GPA)
1	500.00000	661.00000	2.30000
2	460.00000	692.00000	1.40000
3	717.00000	672.00000	2.80000
4	592.00000	441.00000	2.40000
5	752.00000	729.00000	3.40000
6	695.00000	681.00000	2.50000
7	610.00000	777.00000	3.60000
8	620.00000	638.00000	2.60000
9	682.00000	701.00000	3.60000
10	524.00000	700.00000	2.90000
11	552.00000	692.00000	2.60000
12	703.00000	710.00000	3.80000
13	584.00000	738.00000	3.00000
14	550.00000	638.00000	2.50000
15	659.00000	672.00000	3.50000
16	585.00000	605.00000	2.00000
17	578.00000	614.00000	3.00000
18	533.00000	630.00000	2.00000
19	532.00000	586.00000	1.80000
20	708.00000	701.00000	2.30000
21	537.00000	681.00000	2.10000
22	635.00000	647.00000	3.00000
23	591.00000	614.00000	3.30000
24	552.00000	669.00000	3.00000
25	557.00000	674.00000	3.20000
26	599.00000	664.00000	2.30000
27	540.00000	658.00000	3.30000
28	752.00000	737.00000	3.30000
29	726.00000	800.00000	3.90000
30	630.00000	668.00000	2.10000
31	558.00000	567.00000	2.60000
32	646.00000	771.00000	2.40000
33	643.00000	719.00000	3.30000
34	606.00000	755.00000	3.10000
35	682.00000	652.00000	3.60000
36	565.00000	672.00000	2.90000
37	578.00000	629.00000	2.40000
38	488.00000	611.00000	1.80000
39	361.00000	602.00000	2.40000
40	560.00000	639.00000	2.90000
41	630.00000	647.00000	3.50000
42	666.00000	705.00000	3.40000
43	719.00000	668.00000	2.30000
44	669.00000	701.00000	2.90000
45	571.00000	647.00000	1.80000
46	520.00000	583.00000	2.80000
47	571.00000	593.00000	2.30000
48	539.00000	601.00000	2.50000
49	580.00000	630.00000	2.40000
50	629.00000	695.00000	2.90000
51	623.00000	509.00000	2.60000
52	454.00000	471.00000	2.30000
53	643.00000	700.00000	2.40000
54	585.00000	719.00000	3.00000
55	719.00000	710.00000	3.10000
56	693.00000	643.00000	2.90000
57	571.00000	665.00000	3.10000

58	646.00000	719.00000	3.30000
59	613.00000	693.00000	2.30000
60	655.00000	701.00000	3.30000
61	662.00000	614.00000	2.60000
62	585.00000	557.00000	3.30000
63	580.00000	611.00000	2.00000
64	648.00000	701.00000	3.00000
65	405.00000	611.00000	1.90000
66	506.00000	681.00000	2.70000
67	669.00000	653.00000	2.00000
68	558.00000	500.00000	3.30000
69	577.00000	635.00000	2.00000
70	487.00000	584.00000	2.30000
71	682.00000	629.00000	3.30000
72	565.00000	624.00000	2.80000
73	552.00000	665.00000	1.70000
74	567.00000	724.00000	2.40000
75	745.00000	746.00000	3.40000
76	610.00000	653.00000	2.80000
77	493.00000	605.00000	2.40000
78	571.00000	566.00000	1.90000
79	682.00000	724.00000	2.50000
80	600.00000	677.00000	2.30000
81	740.00000	729.00000	3.40000
82	593.00000	611.00000	2.80000
83	488.00000	683.00000	1.90000
84	526.00000	777.00000	3.00000
85	630.00000	605.00000	3.70000
86	586.00000	653.00000	2.30000
87	610.00000	674.00000	2.90000
88	695.00000	634.00000	3.30000
89	539.00000	601.00000	2.10000
90	490.00000	701.00000	1.20000
91	509.00000	547.00000	3.30000
92	667.00000	753.00000	2.00000
93	597.00000	652.00000	3.10000
94	662.00000	664.00000	2.60000
95	566.00000	664.00000	2.40000
96	597.00000	602.00000	2.40000
97	604.00000	557.00000	2.30000
98	519.00000	529.00000	3.00000
99	643.00000	715.00000	2.90000
100	606.00000	593.00000	3.40000

Option number = ?

0 — Exit List routine

SELECT ANY KEY

RECODE ROUTINE

Select Special Function Key labeled-RECODE

Option number = ?

2 — Recoding using contiguous unequal intervals is chosen.

Store recoded data in Variable # (<= 4)
?

4 — Recoded data stored in variable 4.

Variable name (<= 10 characters) = ?

RANKS — Variable name or label.

Number of the variable to be recoded = ?

3 — Recode based on variable 3 (GPA).

Number of recoding intervals to be specified (<=20) = ?

```
4                                                          Four intervals
Lower limit of first interval = ?

1.0                                                        See table below for summary of recoded
Upper limit of interval # 1  =                             specifications.
?

2.0
For data falling in interval 1 = [ 1 , 2 ), code =
?

1
Upper limit of interval # 2  =
?

3
For data falling in interval 2 = [ 2 , 3 ), code =
?

2
Upper limit of interval # 3  =
?

3.5
For data falling in interval 3 = [ 3 , 3.5 ), code =
?

3
Upper limit of interval # 4  =
?

4
For data falling in interval 4 = [ 3.5 , 4 ), code =
?

4
Is above information correct?

YES

Variable # 3 is recoded into 4 categories, and the recoded
values are stored in Variable # 4 , where:
```

Summary: Note that upper limit is not closed but open. That is a value of 3.5 would be recoded as a 4.

CATEGORY BOUNDS LOWER	UPPER	# OBS CODED	CODE
1.000	2.000	9	1.000
2.000	3.000	54	2.000
3.000	3.500	29	3.000
3.500	4.000	8	4.000

```
Option number = ?

0                                                          Exit Recode routine.
PROGRAM NOW UPDATING SCRATCH DATA FILE
SELECT ANY KEY
```

LIST ROUTINE

Select Special Function Key labeled-LIST

```
Option number = ?

1                                                          List all the data.
Enter method for listing data:
3                                                          In tabular form.
```

TOTAL ACT SCORE/GPA COMPARISON DATA

Data type is: Raw data

OBS#	Variable # 1 (VERB)	Variable # 2 (MATH)	Variable # 3 (GPA)	Variable # 4 (RANKS)
1	500.00000	661.00000	2.30000	2.00000
2	460.00000	692.00000	1.40000	1.00000
3	717.00000	672.00000	2.80000	2.00000
4	592.00000	441.00000	2.40000	2.00000
5	752.00000	729.00000	3.40000	3.00000
6	695.00000	681.00000	2.50000	2.00000
7	610.00000	777.00000	3.60000	4.00000
8	620.00000	638.00000	2.60000	2.00000
9	682.00000	701.00000	3.60000	4.00000
10	524.00000	700.00000	2.90000	2.00000
11	552.00000	692.00000	2.60000	2.00000
12	703.00000	710.00000	3.80000	4.00000
13	584.00000	738.00000	3.00000	3.00000
14	550.00000	638.00000	2.50000	2.00000
15	659.00000	672.00000	3.50000	4.00000
16	585.00000	605.00000	2.00000	2.00000
17	578.00000	614.00000	3.00000	3.00000
18	533.00000	630.00000	2.00000	2.00000
19	532.00000	586.00000	1.80000	1.00000
20	708.00000	701.00000	2.30000	2.00000
21	537.00000	681.00000	2.10000	2.00000
22	635.00000	647.00000	3.00000	3.00000
23	591.00000	614.00000	3.30000	3.00000
24	552.00000	669.00000	3.00000	3.00000
25	557.00000	674.00000	3.20000	3.00000
26	599.00000	664.00000	2.30000	2.00000
27	540.00000	658.00000	3.30000	3.00000
28	752.00000	737.00000	3.30000	3.00000
29	726.00000	800.00000	3.90000	4.00000
30	630.00000	668.00000	2.10000	2.00000
31	558.00000	567.00000	2.60000	2.00000
32	646.00000	771.00000	2.40000	2.00000
33	643.00000	719.00000	3.30000	3.00000
34	606.00000	755.00000	3.10000	3.00000
35	682.00000	652.00000	3.60000	4.00000
36	565.00000	672.00000	2.90000	2.00000
37	578.00000	629.00000	2.40000	2.00000
38	488.00000	611.00000	1.80000	1.00000
39	361.00000	602.00000	2.40000	2.00000
40	560.00000	639.00000	2.90000	2.00000
41	630.00000	647.00000	3.50000	4.00000
42	666.00000	705.00000	3.40000	3.00000
43	719.00000	668.00000	2.30000	2.00000
44	669.00000	701.00000	2.90000	2.00000
45	571.00000	647.00000	1.80000	1.00000
46	520.00000	583.00000	2.80000	2.00000
47	571.00000	593.00000	2.30000	2.00000
48	539.00000	601.00000	2.50000	2.00000
49	580.00000	630.00000	2.40000	2.00000
50	629.00000	695.00000	2.90000	2.00000
51	623.00000	509.00000	2.60000	2.00000
52	454.00000	471.00000	2.30000	2.00000
53	643.00000	700.00000	2.40000	2.00000
54	585.00000	719.00000	3.00000	3.00000
55	719.00000	710.00000	3.10000	3.00000
56	693.00000	643.00000	2.90000	2.00000
57	571.00000	665.00000	3.10000	3.00000
58	646.00000	719.00000	3.30000	3.00000

59	613.00000	693.00000	2.30000	2.00000
60	655.00000	701.00000	3.30000	3.00000
61	662.00000	614.00000	2.60000	2.00000
62	585.00000	557.00000	3.30000	3.00000
63	580.00000	611.00000	2.00000	2.00000
64	648.00000	701.00000	3.00000	3.00000
65	405.00000	611.00000	1.90000	1.00000
66	506.00000	681.00000	2.70000	2.00000
67	669.00000	653.00000	2.00000	2.00000
68	558.00000	500.00000	3.30000	3.00000
69	577.00000	635.00000	2.00000	2.00000
70	487.00000	584.00000	2.30000	2.00000
71	682.00000	629.00000	3.30000	3.00000
72	565.00000	624.00000	2.80000	2.00000
73	552.00000	665.00000	1.70000	1.00000
74	567.00000	724.00000	2.40000	2.00000
75	745.00000	746.00000	3.40000	3.00000
76	610.00000	653.00000	2.80000	2.00000
77	493.00000	605.00000	2.40000	2.00000
78	571.00000	566.00000	1.90000	1.00000
79	682.00000	724.00000	2.50000	2.00000
80	600.00000	677.00000	2.30000	2.00000
81	740.00000	729.00000	3.40000	3.00000
82	593.00000	611.00000	2.80000	2.00000
83	488.00000	683.00000	1.90000	1.00000
84	526.00000	777.00000	3.00000	3.00000
85	630.00000	605.00000	3.70000	4.00000
86	586.00000	653.00000	2.30000	2.00000
87	610.00000	674.00000	2.90000	2.00000
88	695.00000	634.00000	3.30000	3.00000
89	539.00000	601.00000	2.10000	2.00000
90	490.00000	701.00000	1.20000	1.00000
91	509.00000	547.00000	3.30000	3.00000
92	667.00000	753.00000	2.00000	2.00000
93	597.00000	652.00000	3.10000	3.00000
94	662.00000	664.00000	2.60000	2.00000
95	566.00000	664.00000	2.40000	2.00000
96	597.00000	602.00000	2.40000	2.00000
97	604.00000	557.00000	2.30000	2.00000
98	519.00000	529.00000	3.00000	3.00000
99	643.00000	715.00000	2.90000	2.00000
100	606.00000	593.00000	3.40000	3.00000

```
Option number =

0                                                          Exit List routine
SELECT ANY KEY
                                   SUBFILE ROUTINES
                                                           Select Special Function Key labeled-SUBFILES
Option number = ?
                                                           Choose to create subfile by values of a
3                                                          variable.
Which variable should be used to create the subfiles ?
                                                           Enter variable no. to be used in creating
4                                                          subfiles.
Criterion value = 1      Enter name for subfile 1 (<=10 characters)
?
POOR
Criterion value = 2      Enter name for subfile 2 (<=10 characters)
?
AVERAGE
Criterion value = 3      Enter name for subfile 3 (<=10 characters)
?
GOOD
Criterion value = 4      Enter name for subfile 4 (<=10 characters)
```

```
?
EXCELLENT
Is the above information correct ?
YES
Subfile name:     beginning observation--number of observations
 1. POOR                                  1                    9
 2. AVERAGE                              10                   54
 3. GOOD                                 64                   29
 4. EXCELLENT                            93                    8

Option number = ?
```

```
0
PROGRAM NOW STORING DATA
SELECT ANY KEY
```

Exit Subfile routine

LIST ROUTINE

Select Special Function Key labeled-LIST

```
Option number = ?
```

```
1
Enter method for listing data:
```

List all the data

```
3
```

In tabular form

```
TOTAL ACT SCORE/GPA COMPARISON DATA

Data type is: Raw data
```

Data is again listed but has now been rearranged on the basis of variable 4.

OBS#	Variable # 1 (VERB)	Variable # 2 (MATH)	Variable # 3 (GPA)	Variable # 4 (RANKS)
1	460.00000	692.00000	1.40000	1.00000
2	532.00000	586.00000	1.60000	1.00000
3	488.00000	611.00000	1.80000	1.00000
4	571.00000	647.00000	1.80000	1.00000
5	405.00000	611.00000	1.90000	1.00000
6	552.00000	665.00000	1.70000	1.00000
7	571.00000	566.00000	1.90000	1.00000
8	488.00000	683.00000	1.90000	1.00000
9	490.00000	701.00000	1.20000	1.00000
10	500.00000	661.00000	2.30000	2.00000
11	717.00000	672.00000	2.80000	2.00000
12	592.00000	441.00000	2.40000	2.00000
13	695.00000	681.00000	2.50000	2.00000
14	620.00000	638.00000	2.60000	2.00000
15	524.00000	700.00000	2.90000	2.00000
16	552.00000	692.00000	2.60000	2.00000
17	550.00000	638.00000	2.50000	2.00000
18	585.00000	605.00000	2.00000	2.00000
19	533.00000	630.00000	2.00000	2.00000
20	708.00000	701.00000	2.30000	2.00000
21	537.00000	681.00000	2.10000	2.00000
22	599.00000	664.00000	2.30000	2.00000
23	630.00000	668.00000	2.10000	2.00000
24	558.00000	567.00000	2.60000	2.00000
25	646.00000	771.00000	2.40000	2.00000
26	565.00000	672.00000	2.90000	2.00000
27	578.00000	629.00000	2.40000	2.00000
28	361.00000	602.00000	2.40000	2.00000
29	560.00000	639.00000	2.90000	2.00000
30	719.00000	668.00000	2.30000	2.00000
31	669.00000	701.00000	2.90000	2.00000
32	520.00000	583.00000	2.80000	2.00000
33	571.00000	593.00000	2.30000	2.00000
34	539.00000	601.00000	2.50000	2.00000

35	580.00000	630.00000	2.40000	2.00000
36	629.00000	695.00000	2.90000	2.00000
37	623.00000	509.00000	2.60000	2.00000
38	454.00000	471.00000	2.30000	2.00000
39	643.00000	700.00000	2.40000	2.00000
40	693.00000	643.00000	2.90000	2.00000
41	613.00000	693.00000	2.30000	2.00000
42	662.00000	614.00000	2.60000	2.00000
43	580.00000	611.00000	2.00000	2.00000
44	506.00000	681.00000	2.70000	2.00000
45	669.00000	653.00000	2.00000	2.00000
46	577.00000	635.00000	2.00000	2.00000
47	487.00000	584.00000	2.30000	2.00000
48	565.00000	624.00000	2.80000	2.00000
49	567.00000	724.00000	2.40000	2.00000
50	610.00000	653.00000	2.80000	2.00000
51	493.00000	605.00000	2.40000	2.00000
52	682.00000	724.00000	2.50000	2.00000
53	600.00000	677.00000	2.30000	2.00000
54	593.00000	611.00000	2.80000	2.00000
55	586.00000	653.00000	2.30000	2.00000
56	610.00000	674.00000	2.90000	2.00000
57	539.00000	601.00000	2.10000	2.00000
58	667.00000	753.00000	2.00000	2.00000
59	662.00000	664.00000	2.60000	2.00000
60	566.00000	664.00000	2.40000	2.00000
61	597.00000	602.00000	2.40000	2.00000
62	604.00000	557.00000	2.30000	2.00000
63	643.00000	715.00000	2.90000	2.00000
64	752.00000	729.00000	3.40000	3.00000
65	584.00000	738.00000	3.00000	3.00000
66	578.00000	614.00000	3.00000	3.00000
67	635.00000	647.00000	3.00000	3.00000
68	591.00000	614.00000	3.30000	3.00000
69	552.00000	669.00000	3.00000	3.00000
70	557.00000	674.00000	3.20000	3.00000
71	540.00000	658.00000	3.30000	3.00000
72	752.00000	737.00000	3.30000	3.00000
73	643.00000	719.00000	3.30000	3.00000
74	606.00000	755.00000	3.10000	3.00000
75	666.00000	705.00000	3.40000	3.00000
76	585.00000	719.00000	3.00000	3.00000
77	719.00000	710.00000	3.10000	3.00000
78	571.00000	665.00000	3.10000	3.00000
79	646.00000	719.00000	3.30000	3.00000
80	655.00000	701.00000	3.30000	3.00000
81	585.00000	557.00000	3.30000	3.00000
82	648.00000	701.00000	3.00000	3.00000
83	558.00000	500.00000	3.30000	3.00000
84	682.00000	629.00000	3.30000	3.00000
85	745.00000	746.00000	3.40000	3.00000
86	740.00000	729.00000	3.40000	3.00000
87	526.00000	777.00000	3.00000	3.00000
88	695.00000	634.00000	3.30000	3.00000
89	509.00000	547.00000	3.30000	3.00000
90	597.00000	652.00000	3.10000	3.00000
91	519.00000	529.00000	3.00000	3.00000
92	606.00000	593.00000	3.40000	3.00000
93	610.00000	777.00000	3.60000	4.00000
94	682.00000	701.00000	3.60000	4.00000
95	703.00000	710.00000	3.80000	4.00000
96	659.00000	672.00000	3.50000	4.00000
97	726.00000	800.00000	3.90000	4.00000
98	682.00000	652.00000	3.60000	4.00000
99	630.00000	647.00000	3.50000	4.00000
100	630.00000	605.00000	3.70000	4.00000

```
Option number = ?

0                                                              Exit List routine
SELECT ANY KEY                          STORE ROUTINE
                                                               Select Special Function Key labeled-STORE
Enter option number desired :

1                                                              Store the complete set of data.
Name of data file = ?

TGRADE:INTERNAL                                                On this file.
Is data medium placed in device ?
?

YES
PROGRAM NOW STORING DATA ON TGRADE:INTERNAL



* * * * * The data and related information are stored in TGRADE:INTERNAL * * * * *


Is program medium placed in device ?

YES
Enter option number desired :

0                                                              Exit Store routine.
SELECT ANY KEY
                                        SELECTION ROUTINES
                                                               Choose Special Function Key labeled-SELECT
Choose option desired :

2                                                              Select choosen instead of Scan.
Choose option desired :

1                                                              Choose to Select on basis of value of just
SELECTION BASED ON ONE VARIABLE                                one variable.
Which variable should be used ?

1                                                              Variable 1 = Verb
Criterion variable = 1    (VERB)
What values can the criterion variable take ?

550-800                                                        Select those cases for which Verb is be-
Allowable values : 550-800                                     tween 550 and 800.
Which subfiles do you want to be selected ?


ALL                                                            For both subfiles.
SUBFILES TO BE SELECTED : ALL


OBSERVATIONS SATISFYING SELECTION CRITERION :
   3     5     9    10    12    14    15    16    17    19
  20    22    23    24    25    26    28    29    30    31
  32    33    35    36    37    38    40    42    43    44
  45    47    49    50    51    53    54    55    56    57        These observations
  58    59    61    62    63    64    65    67    68    69        meet the criteria.
  70    71    72    73    74    75    76    78    79    80
  81    82    84    85    86    87    88    89    90    91
  93    94    95    96    97    98    99   100
```

```
                  BEFORE SELECTION   AFTER SELECTION
SUBFILE             NUM OF OBS         NUM OF OBS

POOR                    9                  3
AVERAGE                54                 42
GOOD                   29                 25
EXCELLENT               8                  8
PROGRAM NOW UPDATING SCRATCH DATA FILE
Choose option desired :
```

The Selection routine saves only those observations whose verbal score was between 550 - 800. The rest of the observations are discarded from the program memory.

```
0
SELECT ANY KEY
```

Exit Select routine.

STATS ROUTINE

Select Special Function Key labeled-STATS

```
What statistic options are desired ?

1
VARIABLES=
?
```

Mean, CI, Variance, Standard Deviation, Skewness, Kurtosis.

```
ALL
Confidence coefficient for confidence interval on the mean(e.g. 90,95,99%) =
```

Statistics will be computed for all variables.

```
95
Option number = ?
```

With a 95% coefficient.

```
2
What subfiles are desired ?
1-4
```

Complete statistics for specified subfiles.

All subfiles

```
******************************************************************************
*                          SUMMARY STATISTICS                                *
*                             ON DATA SET:                                   *
*                 TOTAL ACT SCORE/GPA COMPARISON DATA                        *
******************************************************************************
------------------------------------------------------------------------------
Subfile: POOR
------------------------------------------------------------------------------
```

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
VERB	3	0	1694.0000	564.6667	120.3333	10.9697
MATH	3	0	1878.0000	626.0000	2781.0000	52.7352
GPA	3	0	5.4000	1.8000	.0100	.1000
RANKS	3	0	3.0000	1.0000	0.0000	0.0000

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
VERB	1.94268	6.33333	537.60540	591.72793
MATH	8.42415	30.44667	495.90649	756.09351
GPA	5.55556	.05774	1.55331	2.04669
RANKS	0.00000	0.00000	1.00000	1.00000

VARIABLE	SKEWNESS	KURTOSIS
VERB	-.70711	-1.50000
MATH	-.61556	-1.50000
GPA	0.00000	-1.50000
RANKS	-------------	-------------

Subfile: AVERAGE

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
VERB	42	0	25935.0000	617.5000	2382.4024	48.8099
MATH	42	0	27318.0000	650.4286	3694.4460	60.7820
GPA	42	0	104.3000	2.4833	.0814	.2853
RANKS	42	0	84.0000	2.0000	0.0000	0.0000

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
VERB	7.90443	7.53152	602.28627	632.71373
MATH	9.34491	9.37886	631.48322	669.37393
GPA	11.49047	.04403	2.39439	2.57227
RANKS	0.00000	0.00000	2.00000	2.00000

VARIABLE	SKEWNESS	KURTOSIS
VERB	.54518	-.82101
MATH	-1.03447	2.32038
GPA	-.03388	-.90383
RANKS	-------------	-------------

Subfile: GOOD

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
VERB	25	0	15948.0000	637.9200	4324.1600	65.7583
MATH	25	0	16856.0000	674.2400	4096.6067	64.0047
GPA	25	0	80.3000	3.2120	.0236	.1536
RANKS	25	0	75.0000	3.0000	0.0000	0.0000

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
VERB	10.30824	13.15167	610.76982	665.07018
MATH	9.49287	12.80095	647.81385	700.66615
GPA	4.78278	.03072	3.14857	3.27543
RANKS	0.00000	0.00000	3.00000	3.00000

VARIABLE	SKEWNESS	KURTOSIS
VERB	.48079	-1.04529
MATH	-.96523	.42114
GPA	-.27487	-1.47768
RANKS	-------------	-------------

Subfile: EXCELLENT

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
VERB	8	0	5322.0000	665.2500	1607.6429	40.0954
MATH	8	0	5564.0000	695.5000	4398.5714	66.3217
GPA	8	0	29.2000	3.6500	.0200	.1414
RANKS	8	0	32.0000	4.0000	0.0000	0.0000

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
VERB	6.02712	14.17587	631.72037	698.77963
MATH	9.53583	23.44827	640.03874	750.96126
GPA	3.87456	.05000	3.53174	3.76826
RANKS	0.00000	0.00000	4.00000	4.00000

VARIABLE	SKEWNESS	KURTOSIS
VERB	.07320	-1.21757
MATH	.38485	-.97545
GPA	.64794	-.77551
RANKS	-------------	-------------

```
What statistic options are desired ?
2                                          Correlation matrix
VARIABLES=
?
ALL                                        Statistics completed for all variables
Option number = ?
2                                          Compute statistics for specified subfiles
What subfiles are desired ?
1-4                                        All subfiles
```

```
******************************************************************************
*                           SUMMARY STATISTICS                               *
*                              ON DATA SET:                                  *
*                  TOTAL ACT SCORE/GPA COMPARISON DATA                       *
******************************************************************************
------------------------------------------------------------------------------
Subfile: POOR
------------------------------------------------------------------------------
                           CORRELATION MATRIX


                   MATH        GPA      RANKS
VERB           -.6404640   .8660254  ----------
MATH                      -.9386522  ----------
GPA                                  ----------



------------------------------------------------------------------------------
------------------------------------------------------------------------------
Subfile: AVERAGE
------------------------------------------------------------------------------
                           CORRELATION MATRIX


                   MATH        GPA      RANKS
VERB            .3530502   .0440427  ----------
MATH                       .0482350  ----------
GPA                                  ----------



------------------------------------------------------------------------------
------------------------------------------------------------------------------
Subfile: GOOD
------------------------------------------------------------------------------
                           CORRELATION MATRIX


                   MATH        GPA      RANKS
VERB            .4981619   .5173239  ----------
MATH                      -.0706494  ----------
GPA                                  ----------



------------------------------------------------------------------------------
------------------------------------------------------------------------------
Subfile: EXCELLENT
------------------------------------------------------------------------------
                           CORRELATION MATRIX


                   MATH        GPA      RANKS
VERB            .3654701   .6651140  ----------
MATH                       .4934875  ----------
GPA                                  ----------
------------------------------------------------------------------------------
What statistic options are desired ?
3
VARIABLES=
?
ALL
Option number = ?
2
What subfiles are desired ?
1-4
```

Median mode, percentiles, Min., Max., Range

Statistics computed for all variables

Compute Statistics for specified subfiles

All subfiles

```
****************************************************************************
*                          SUMMARY STATISTICS                              *
*                             ON DATA SET:                                 *
*                 TOTAL ACT SCORE/GPA COMPARISON DATA                      *
****************************************************************************
----------------------------------------------------------------------------
Subfile: POOR
----------------------------------------------------------------------------


                           ORDER STATISTICS


VARIABLE            MAXIMUM          MINIMUM          RANGE         MIDRANGE
VERB              571.00000        552.00000       19.00000        561.50000
MATH              665.00000        566.00000       99.00000        615.50000
GPA                 1.90000          1.70000         .20000          1.80000
RANKS               1.00000          1.00000        0.00000          1.00000
----------------------------------------------------------------------------
                                             TUKEY'S HINGES

VARIABLE                MEDIAN      25-th %-ile      75-th %-ile
VERB                 571.00000        552.00000        571.00000
MATH                 647.00000        566.00000        647.00000
GPA                    1.80000          1.70000          1.80000
RANKS                  1.00000          1.00000          1.00000
----------------------------------------------------------------------------
                                 TUKEY'S MIDDLEMEANS
VARIABLE
                       MIDMEAN          TRIMEAN        MIDSPREAD
VERB                 564.66667        566.25000         19.00000
MATH                 626.00000        626.75000         81.00000
GPA                    1.80000          1.77500           .10000
RANKS                  1.00000          1.00000          0.00000
----------------------------------------------------------------------------
Other percentiles(Y/N)?
NO
----------------------------------------------------------------------------
----------------------------------------------------------------------------
Subfile: AVERAGE
----------------------------------------------------------------------------


                           ORDER STATISTICS


VARIABLE            MAXIMUM          MINIMUM          RANGE         MIDRANGE
VERB              719.00000        550.00000      169.00000        634.50000
MATH              771.00000        441.00000      330.00000        606.00000
GPA                 2.90000          2.00000         .90000          2.45000
RANKS               2.00000          2.00000        0.00000          2.00000
----------------------------------------------------------------------------
                                             TUKEY'S HINGES

VARIABLE                MEDIAN      25-th %-ile      75-th %-ile
VERB                 607.00000        578.00000        646.00000
MATH                 658.50000        624.00000        681.00000
GPA                    2.40000          2.30000          2.60000
RANKS                  2.00000          2.00000          2.00000
----------------------------------------------------------------------------
```

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
VERB	610.13636	609.50000	68.00000
MATH	655.95455	655.50000	57.00000
GPA	2.46818	2.42500	.30000
RANKS	2.00000	2.00000	0.00000

Other percentiles(Y/N)?
NO

Subfile: GOOD

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
VERB	752.00000	552.00000	200.00000	652.00000
MATH	755.00000	500.00000	255.00000	627.50000
GPA	3.40000	3.00000	.40000	3.20000
RANKS	3.00000	3.00000	0.00000	3.00000

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
VERB	635.00000	585.00000	666.00000
MATH	701.00000	634.00000	719.00000
GPA	3.30000	3.10000	3.30000
RANKS	3.00000	3.00000	3.00000

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
VERB	626.53846	630.25000	81.00000
MATH	685.76923	688.75000	85.00000
GPA	3.23077	3.25000	.20000
RANKS	3.00000	3.00000	0.00000

Other percentiles(Y/N)?
NO

Subfile: EXCELLENT

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
VERB	726.00000	610.00000	116.00000	668.00000
MATH	800.00000	605.00000	195.00000	702.50000
GPA	3.90000	3.50000	.40000	3.70000
RANKS	4.00000	4.00000	0.00000	4.00000

```
                                          TUKEY'S HINGES

VARIABLE                 MEDIAN          25-th %-ile          75-th %-ile
VERB                  670.50000            630.00000            692.50000
MATH                  686.50000            649.50000            743.50000
GPA                     3.60000              3.55000              3.75000
RANKS                   4.00000              4.00000              4.00000
--------------------------------------------------------------------------------
                                 TUKEY'S MIDDLEMEANS
VARIABLE
                        MIDMEAN              TRIMEAN            MIDSPREAD
VERB                  663.25000            665.87500             62.50000
MATH                  683.75000            691.50000             94.00000
GPA                     3.62500              3.62500               .20000
RANKS                   4.00000              4.00000              0.00000
--------------------------------------------------------------------------------
Other percentiles(Y/N)?
NO
--------------------------------------------------------------------------------
What statistic options are desired ?
0
SELECT ANY KEY
```

Exit Basic Statistics routine

Regression Analysis

General Information

Description

The Regression Analysis software provides you with five routines to perform various types of linear and non-linear regressions. The regression routines include:

- Multiple Linear Regression
- Polynomial Regression
- Variable Selection Procedures (Stepwise algorithm, etc.)
- Non-linear Regression
- Standard Non-linear Regression Models

In addition, a residual analysis module is included which will be helpful in judging the quality of the chosen regression model. Brief desciptions of each regression routine follow.

The multiple linear regression routine performs a least-squares regression on a set of predetermined variables.

The variable selection procedures perform least-square regressions iteratively on sets of variables which are determined by one of four selection procedures – stepwise, forward selection, backward elimination, or manual. These selection procedures are helpful in determining which of the independent variables are "important" in predicting the behavior of the dependent variable.

The polynomial regression routine is a special case of the multiple linear regression procedure where the independent variables are actually powers of a single variable. In other words, the form of the regression model is:

$$Y = B0 + B1*(X) + B2*(X \uparrow 2) + ... + Bp*(X \uparrow p),$$

where Y is the dependent variable, X is the independent variable, and B1, ..., Bp are the regression coefficients. A routine is also provided so you can plot the X-Y data along with the regression curve.

The non-linear regression routine allows you to determine the coefficients of virtually any model you wish to specify. It is more difficult to use than the multiple linear regression routines; however, its use is mandatory when the model is non-linear in the regression coefficients. An example of this is the model:

$$Y = B1(Exp)B2*X1 + B3*X2),$$

where Exp is the exponential function. A plotting routine is provided so you can plot any variable versus the dependent variable. If the model has only one independent variable, the regression curve can also be plotted.

The routines referred to as "standard" non-linear regressions determine the regression coefficients for the following four types of common non-linear regression models:

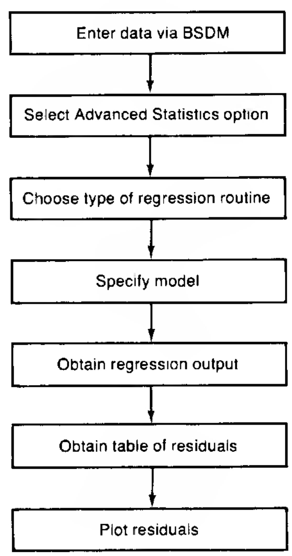
- Y = A*X ↑ B + C
- Y = A*Exp(BX) + C
- Y = A*Exp(BX) + C*Exp(DX) + E
- Y = A*Sin(BX) + C*Cos(DX) + E

Also provided is a routine to plot the data along with the computed regression curve.

All of the regression programs provide an analysis of variance table, correlations, and the regression coefficients, as well as their standard errors.

The residual analysis routine provides a list of the residuals as well as a plot of the standardized residuals versus observation number or any variable.

Typical Program Flow



Special Considerations

Terminology

By an independent variable we mean a variable that can be set to a desired value (for example, input temperature or catalyst feed rate in a chemical reaction), or values that can be observed but not controlled (for example, the outdoor humidity).

As a result of changes in one or more independent variables, the dependent variable will be affected. For example, the purity of a chemical product may be affected by temperature and the catalyst feed rate.

In a simple linear regression: Y = B0 + B1*X, Y is the dependent variable, and X is the independent variable, while B0 and B1 are the regression coefficients.

Data Structure

Data is input via the Basic Statistics and Data Manipulation routines. You need to tell the regression routine the number of the BSDM variable which you want to be your dependent variable. In general, you tell the routine how many independent variables are in your regression model. Then, you specify the BSDM variable numbers which you want to be your independent variables. For example, suppose you input 10 variables in the BSDM procedure. You might specify that variable #4 is your dependent variable and that you want to have five independent variables. You then might specify the independent variables as BSDM variables #2, #3, #5, #7, and #9.

If you specify subfiles with the BSDM procedure, you may perform regressions on individual subfiles.

Note

Non-Linear Regression

You will have to create a file which contains the function and partial derivatives before you get into the program. The steps involved are shown on page 69.

Multiple Linear Regression

Object of Program

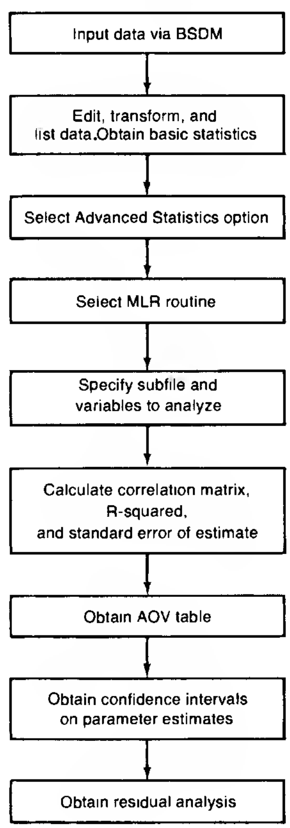
This routine is designed to calculate a least-squares multiple linear regression on a predetermined set of variables. The general form of the regression model is:

$$Y = B0 + B1X1 + B2X2 + \ldots + BpXp + \text{Error}$$

where Y is the dependent variable, X1, X2, ..., Xp are the independent variables and B0, B1, ..., Bp are the regression coefficients.

Several basic statistics, as well as the correlation matrix, are output. An analysis of variance table is printed. The regression coefficients and their standard errors are output and confidence intervals are constructed about them. Output along with each regression coefficient is an associated t-value. This statistic is used to test if the regression coefficient is significantly different from zero, i.e., if the term is useful in the model. In addition, the regression equation may be used for predictions and a residual analysis may be performed.

Typical Program Flow



Special Considerations

Method of Computing Sums of Squares and Cross Products Matrix

If a data value is missing for one or more variables, the entire observation is deleted, i.e., not used in computing the sums of squares and cross products matrix (or correlations). Consider the following matrix where missing values are denoted by an M.

		Variable 1	Variable 2	Variable 3
Observation	1	M	3	2
	2	1	3	4
	3	2	2	3
	4	M	4	M
	5	1	3	3

Observation 1 is deleted since the data value is missing for variable 1 and observation 4 is deleted since the data value is missing for variables 1 and 3. Hence, only obervations 2, 3, and 5 will be used to compute the sums of squares and cross products matrix, as well as the correlations.

Constant Term

In the output of the regression coefficients, the term labeled "Constant" refers to the intercept or initial value when all the independent variables are zero. This constant term corresponds to the B0 term in the general form of the model shown in the Object of Program section.

Transforming Variables

After you input your data via Basic Statistics and Data Manipulation, you can use the transformation routine to create new variables. The transformation routine has several predefined functions which will allow you to create transgenerated regression variables. Refer to the Basic Statistics and Data Manipulation section for further details on transforming variables.

Additional Sum of Squares in AOV Table

In the analysis of variance table, you will see that the degrees of freedom and the sum of squares of regression are dividied into several parts, each with one degree of freedom. For example, suppose a regression problem has three independent variables, say X1, X2, and X3. You will notice that these three variables are listed below the "regression" term in the AOV table, and that each has one degree of freedom. See the sample problem on page 25.

The meaning for the X1 line is as follows. We first consider only X1 in the regression model and from the sum of squares we can tell how much of the variation of the dependent variable is explained by introducing X1 into the model. The meaning for the X2 line is as follows. Given that X1 is in the model, if we introduce X2 into the model we can see how much additional variation is explained by X2. Then, in the X3 line, we suppose X1 and X2 are already in the model. The sum of squares shows how much additional variation is explained by adding X3 to the regression model. The total degrees of freedom of the independent variables are equal to the regression degrees of freedom. The sum of squares of the independent variables will also add up to the sum of squares for regression.

Methods and Formulae

The Cholesky square-root method is used to factor the sum of squares and cross products matrix. It is felt that this method produces less round off error than other inversion techniques. This method, as well as all other methods and formulae used may be found in F.A. Graybill's Theory and Application of the Linear Model, Chapters 7 and 10.

Stepwise Regression (Variable Selection Procedures)

Object of Program

This program allows a regression model to be built iteratively using one of four variable selection procedures. The procedures are stepwise, forward, backward, and manual. A correlation matrix is calculated and output. An analysis of variance table, as well as partial correlations, F values for deletion and inclusion, and the regression coefficients are output at each step of the regression. In addition, a residual analysis may be performed.

The four selection procedures operate as follows:

Stepwise

You specify an F-to-enter and an F-to-delete, and the program begins with no variables in the regression model. If any of the variables have an F value larger than the F-to-enter, then the variable with the largest F value is entered into the model. This process is repeated with the remaining variables. At this point, the F values of the variables in the model are compared with the F-to-delete. If a variable has a smaller F value than the F-to-delete, it is removed from the model. This process of adding and deleting variables continues until all the variables in the model have F values larger than the F-to-delete and all the variables not in the model have F values smaller than the F-to-enter, or until the tolerance value becomes too small. A small tolerance value signals that the matrix has become unstable.

Forward Selection

You input an F-to-enter. The program operates in the same manner as the stepwise selection procedure, except that variables are not deleted. The process continues until all variables not in the model have F values smaller than the F-to-enter, or until the tolerance value becomes too small.

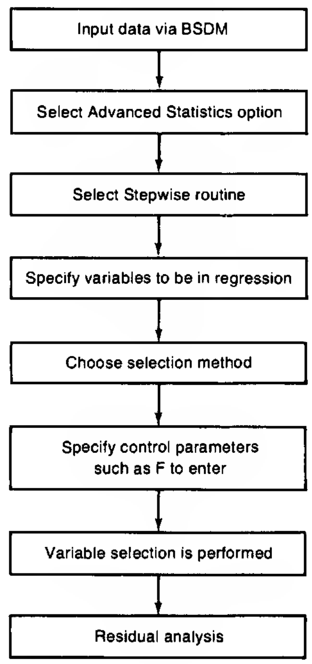
Backward Elimination

You input an F-to-delete and the program begins with all the variables in the model. If any variable has an F value smaller than the F-to-delete, then that variable with the smallest F value is deleted from the model. This process continues until all the variables in the model have F values larger than the F-to-delete or until the tolerance value becomes too small.

Manual Selection

As the name implies, variables are added or deleted manually until you are satisfied with the model.

Typical Program Flow



Special Considerations

F Values Insufficient for Further Computation

If one of the stepwise, forward, or backward procedures is used in the selection of variables, the program will proceed automatically by entering and/or removing variables from the model until the F values are not exceeded or until the tolerance value is not met. At this point the program reverts to the manual mode. So, for example, this allows you to enter a variable whose F value is just slightly less than the specified F-to-enter.

Methods of Computing Correlations

Two methods of computing correlations are available. The first method will use an observation only if data values are present for each variable. The second method uses all possible data values to compute each correlation. If no missing values are present, method two should be used to speed computation.

A simple example will show the difference between the two methods. Suppose we have the following data set:

		Variable		
		1	2	3
Observations	1	2	3	M
	2	3	2	4
	3	1	3	5
	4	M	1	4

If method one is used to compute the correlations, only observations 2 and 3 will be used. Observation 1 will be deleted entirely since the data value is missing for variable 3. Similarly, observation 4 will be deleted entirely since the data value is missing for variable 1.

Conversely, suppose method two is chosen. The correlation between variables 1 and 2 will be computed using the data values of observations 1, 2, and 3. The correlation between variables 1 and 3 will use the data values associated with observations 2 and 3. Similarly, the correlation between variables 2 and 3 will use the data values associated with observations 2, 3, and 4. Hence, data values from a given observation are used if the data points are present for the two variables under consideration.

The observations used to compute AOV table are the same as those used to get the correlations.

F-to-enter, F-to-delete

A variable must have an F value which is greater than the value of F-to-enter for entry into the regression model via the stepwise or forward selection procedures. A typical value is 4. A variable may be deleted from the regression via the stepwise or backward selection procedures only if its F value is less than the value of F-to-delete. When using the stepwise procedure, you must have F-to-enter >= F-to-delete. The F-to-enter should be selected from tabled values for your desired significance level with 1 and n-v degrees of freedom, where n is the number of observations and v is the number of variables in the regression. Since you don't know how many variables will be in the regression a priori, you might guess the number of variables which will end up in the regression for your initial analysis.

Tolerance Value
You will be asked to enter a tolerance value. Your input must be between 0 and 1. The tolerance value is a scaled function of the determinant of the X'X matrix, and is a measure of the stability of the correlation matrix. If a variable not in the equation is linearly dependent on one of more of the variables already in the model, then the correlation matrix will have a determinant of zero. So, if the computed tolerance value gets too small, this might suggest a singular matrix. A suggested value for the tolerance is .01.

Reading the Output
In the algorithm, one variable will be entered or deleted per step. The variables currently included in the regression model are printed on the left side of the table. The variables which are not currently included in the model are printed on the right side of the table.

Partial Correlation
The partial correlations of the variables not currently in the regression equation are output. After a variable, say X1, has been entered into the regression model, the program calculates the partial correlation of the other independent variables with the dependent variable, given that X1 is in the regression model.

Adding One Variable to the Model
If any of the variables has an F value larger than the F-to-enter, then the variable with the largest F will be entered into the model provided that its tolerance value is greater than the user specified tolerance value.

Deleting One Variable from the Model
If any variable currently in the regression equation has an F value smaller than F-to-delete, then the one with the smallest F value will be deleted from the model at that step.

Manual Selection
After you have completed a portion of the program, you will see the prompt "Input 'K', delete '−K' ?". At this point the program is operating in a manual mode. That is, you may add a variable to the regression equation by entering its number, or delete a variable from the equation by entering its number preceeded by a minus sign.

Methods and Formulae
All methods and formulae used in this routine may be found in Statistical Methods for Digital Computers by K. Enslein, et.al.

Polynomial Regression

Object of Program

This program is designed to fit a polynomial regression model of the form:

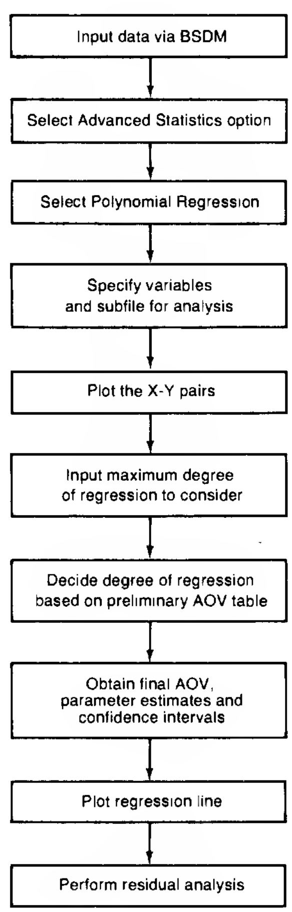
$$Y = B0 + B1(X) + B2(X \uparrow 2) + B3(X \uparrow 3) + \ldots + Bp(X \uparrow p)$$

where $p <= 10$. The regression coefficients, B0, B1, ..., Bp are computed by the method of least squares.

The degree of the regression, p, is chosen by you with the aid of a preliminary analysis of variance table and, if desired, an X-Y scatter plot. The preliminary analysis of variance table shows the additional sum of squares explained by models of successive degrees as well as the associated F values and R-squared values.

After the degree of the regression is selected, an analysis of variance table for the model is printed and confidence intervals are constructed about the coefficients. In addition, a residual analysis may be performed.

Typical Program Flow



Special Considerations

Degree of Model

The maximum degree of the model has been set (somewhat arbitrarily) at 10. Models of degree ten involve arithmetic operations using the X variable raised to the 20th power, where X is the independent variable. Hence, substantial round-off errors may occur with models of high degree. In general, a model of degree p will involve X values raised to the 2*p power. It is therefore suggested that you use extreme caution in choosing models of high degree.

Method of Computing Sums of Squares and Cross Products Matrix

If a data value is missing for one of the two variables, the entire observation is deleted, i.e., not used in the computation of the sums of squares and cross products matrix. See Special Considerations of the Multiple Linear Regression section for an example.

Preliminary AOV Table

After plotting the X-Y data pairs, you will be asked to specify the maximum degree of the regression. A preliminary AOV table will be displayed which will show the additional sum of squares and R-squared for the linear, quadratic, cubic, ... regression models. This table can be used as an aid in determining the appropriate degree for your polynomial model.

Plotting Considerations

When plotting the data and regression, every tic mark on the axes will be labeled. So, you should specify no more than 10 tic marks to obtain an uncluttered plot. One tic mark will coincide with the point where the X-axis crosses the Y-axis. Another tic mark will coincide with the point where the Y-axis crosses the X-axis.

Plotting the data is highly recommended since a plot may suggest the degree of the polynomial model.

Methods and Formulae

The Cholesky square-root method is used to factor the sum of squares and cross products matrix. It is felt that this inversion method produces less round-off error than other procedures. This method, as well as all other methods and formulae may be found if F.A. Graybill's Theory and Application of the Linear Model.

Nonlinear Regression

Object of Program

Given a model

$$Y = f(X_1, X_2 \ldots, Xm; \beta_1 \beta_2, \ldots, \beta p) + \epsilon$$

where the model f contains m independent variables X_i and p parameters βj and given n observations

$$(Yi, Xi_1, Xi_2, \ldots, Xi_m) \quad ; \quad i = 1,2,\ldots,n$$

this program computes the least square estimates $\hat{\beta} j$; that is, the program adjusts the $\hat{\beta} j$ to minimize

$$Q = \sum_{i=1}^{n} \{Yi - f(Xi_1, Xi_2, \ldots, Xim, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta} p)\}^2$$

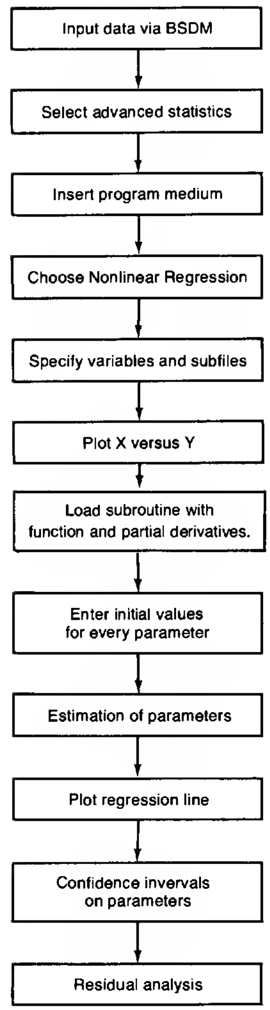
You supply the functional form of f. For example, one possible form would be

$$Y = \beta_1 \exp\ (\beta_2 X_1 + \beta_3 X_2) + \beta_4$$

The program also provides X-Y scatter plots (the non-linear regression curve can be added to the plot if the model contains only one independent variable). After each iteration the following information is output: the iteration number, estimated parameter values, and sum of squared residuals (Q). Confidence intervals (regions) on the parameters are also constructed. In addition, a residual analysis may be performed.

Before beginning the program, you will need to create a file which contains the function and partial derivatives. The necessary steps are shown in the Special Considerations section.

Typical Program Flow



Special Considerations

Limitations

The maximum number of parameters in the model is 20. Also, the number of observations times the number of parameters must be less than or equal to 5000.

Convergence Criteria

From a user viewpoint there are three modes of program termination during the iterative stage of estimation of the parameter. The first mode is the satisfactory completion of the convergence criteria; that is, the iteration is terminated whenever

$$\frac{|\delta j|}{0.001 + |\hat{\beta} j|} < \text{delta for all } j$$

where delta is a small number that you input, and δj is the change in $\hat{\beta} j$ resulting from the last iteration. This is the normal termination which should occur when a proper function has been specified for f, the derivatives are specified correctly, and the initial estimates for the parameters are reasonable.

A second mode of termination can occur when the program determines that the process is not converging in a satisfactory manner. (For the procedure used in determining whether the process is converging properly, see Reference 5.) If the program does terminate the iterative process, you are able to respecify the convergence coefficient (Delta), the function and/or derivatives, and the initial parameter estimates.

The third method of termination of the iterative process is for you to "force off" the computational process by pressing the "No" key.

Quick Plot

A quick plot is essentially a default plot with plotting parameters:

1. X-min = actual X-min, X-max = actual X-max.
2. Y-min = actual Y-min, Y-max = actual Y-max.
3. Y-axis crosses X-axis at X-min.
4. X-axis crosses Y-axis at Y-min.
5. Distance between X-tics = (Xmax-Xmin)/5.
6. Distance between Y-tics = (Ymax-Ymin)/5.
7. Number of decimals for labeling X-axis and Y-axis = 2.

   You may wish to have the quick plot drawn in order to "see" what the relationship between Y and the X you have chosen looks like.

The actual limits of the confidence intervals are very data dependent. Caution should be exercised in using these limits if many iterations were required to determine the regression coefficients.

Before you Run Non-linear Regression

To run non-linear regression, you must first create a file which contains the function and partial derivatives you wish to use. You can create as many of the files as you wish. The procedure to create these files is as follows:

- Insert your floppy in the built-in disc drive
- Type SCRATCH A; press EXECUTE
- Press EDIT key; press EXECUTE
  You should now see the line number ten on the screen.
- Now type in each line of the file, pressing ENTER after every line that has been entered. The file should resemble the one below.

Note

Remember that partial derivatives should be taken with respect to P(*).

```
10   SUB Function(P(*),X(*),F)
20     F=P(1)+P(2)*X(1)^P(3)
30     SUBEND
40   SUB Partial(P(*),X(*),Der(*))
50     Der(1)=1
60     Der(2)=X(1)^P(3)
70     Der(3)=P(2)*LOG(X(1))*X(1)^P(3)
80     SUBEND
```

- The two SUB statements in your file must be exactly the same as in the example.
- When you have finished typing the two subroutines, press the CLR SCR KEY. Type STORE "name of file". You may name your file whatever you like as long as the name is not greater than ten characters long and has nothing but letters and numbers in it.
- You may now begin running the Statistics Library by typing LOAD "AUTOST",1 with the BASIC Statistics and Data Manipulation disc in the internal disc drive.

Methods and Formulae

The Marquardt's procedure (see Reference 5) is used to obtain the estimated parameters in each iteration. Define

$$\underline{Z} = (Zij) = \left[\frac{\partial f(X_1j, X_2j, \ldots, Xmj, \hat{\beta}_1, \ldots \hat{\beta}p)}{\partial \hat{\beta} i} \right] = \left[\frac{(\partial f(Xj, \hat{\beta})}{\partial \hat{\beta} i} \right]$$

then each iteration can be written as

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \delta^{(k)}$$

where $\delta(k)$ is the solution of the set of linear equations

$$(A + \lambda I)\delta = Z'(Y - f(X, \hat{\beta})) = g$$

where $A = Z'Z$ and g are evaluated at $\hat{\beta}(k)$ (both A and g are normalized in the program), and where λ is an adjustable parameter which is used to control the iteration. The motivation of Marquardt's method is to choose λ so as to follow the Gauss-Newton method to as large an extent as possible, while retaining a bias towards the steepest descent direction to prevent divergence.

The square root method is used to solve the system of linear equations in each iteration and to obtain $C = (Cij) = A^{-1}$.

For the confidence intervals (regions) on parameters, the $1 - \alpha$ one-at-a time confidence interval on βj is

$$\hat{\beta} j - t(\alpha/_2 : n - p)(Se^2 Cjj)^{1/2} \leq \beta j \leq \hat{\beta} j + t(\alpha/_2 : n - p)(Se^2 Cjj)^{1/2}$$

and the approximate $1 - \alpha$ simultaneous confidence intervals on βj's are

$$\hat{\beta} j - (pF(\alpha : p, n - p)Se^2 Cjj)^{1/2} \leq \beta j \leq \hat{\beta} j + ((pF(\alpha : p, n - p)Se^2 Cjj)^{1/2}$$

where p is the number of parameters in the model, n is the number of observations (exclude the missing values), $t(\alpha/_2 : n - p)$ is the $\alpha/2$ upper point of the T-distribution with n-p degrees of freedom. $F(\alpha : p, n - p)$ is the α upper point of the F-distribution with p and n-p degrees of freedom, and Se is the standard error of the residuals.

References

1. Draper, N., and Smith, H., (1980) Applied Regression Analysis, 2nd Edition, John Wiley and Sons, Inc., New York.
2. Fletcher, R. (1971) "A Modified Marquardt Subroutine for Nonlinear Least Squares", United Kingdon Atomic Energy Authority Research Group Report.
3. Graybill, F. (1976) Theory and Application of the Linear Model, Wadsworth Publishing Co., Inc., California.
4. Kopitzke, R., and (Boardman, T.J., Editor). Unpublished Notes for 9830A Statistical Distribution Pac. Hewlett-Packard, September 1976. Part No. 09830-70854.
5. Marquardt, D. (1963). "An Algorithm for Least Squares Estimation of Nonlinear Parameters". J. Soc. Indust. and Appl. Math., 11. No. 2.

Standard Nonlinear Regressions

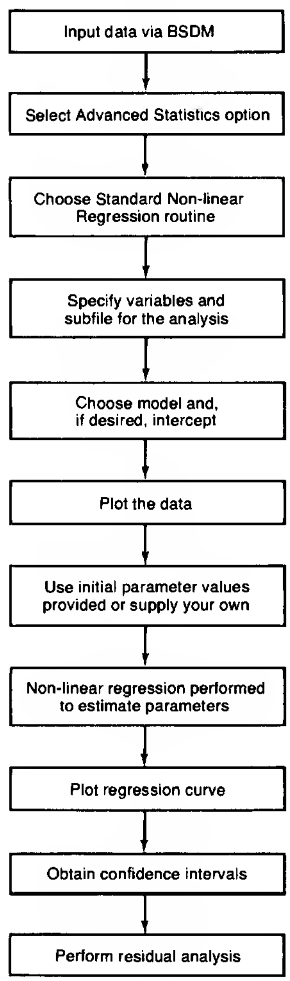
Object of Program

This program determines the regression coefficients for the following four types of standard non-linear regression models:

1. Y = A(X ↑ B) + C
2. Y = A*Exp(BX) + C
3. Y = A*Exp(BX) + C*Exp(DX) + E
4. Y = A*Sin(BX) + C*Cos(DX) + E

where the intercept term, C or E above, is optional. The intercept is determined by using an approximate minimum Y value in the observed data as the initial value.

Typical Program Flow



Special Considerations

Initial Parameter Estimates

In models 1), 2), and 3), initial estimates for parameters are obtained by linearizing the model. This is accomplished by taking the logarithm of both sides of the equation for model 1, and by taking the logarithm of Y in models 2 and 3. In model 3, C is taken as .1*A and D = .5*B. In model 4:

A = (Ymax − E) * Sin(a) * Cos(B * Xmax)
B = 360 / (length in units of X of a typical cycle)
C = (Ymax − E) * Cos(a) * Sin(B*Xmax)
D = B
E = sample mean of y

where a = 90 − B * X1, for data in degrees, and X1 is the X value at Ymax.

For angular units in radians, the estimates of B and C will change accordingly.

Convergence Criteria

There are three ways by which the program may terminate its iterative procedure of estimating the model parameters.

a. The iteration is terminated when

   $|\Delta j| / (.001 + |\hat{\beta} j| <$ Delta for all regression coefficients, $\hat{\beta} j$,

   where Delta is a small number that you input, and Δj is the change in $\hat{\beta} j$ resulting from the last iteration. This is the normal termination which should occur when the proper model has been selected for a given data set and the initial estimates are chosen properly.

b. When the program determines that the process is not converging in a satisfactory manner, it will terminate. For the procedure used in determining whether the procedure is converging properly, see reference 5 in the Non-linear Regression section. If the program does terminate the iterative process, you can re-specify the convergence coefficient (Delta), and/or the initial estimates of the parameters and try the regression again.

c. You may force the iterative procedure to terminate by pressing the "Stop" key.

Angular Units for Model 4

When model 4, the trigonometric model, is chosen, you need to specify two additional items for the program. You must declare whether your X values are in degrees or radians. In addition, during the routine which supplies the initial estimates for the parameters, you need to specify the length of a typical cycle of data.

Residual Analysis

Object of Program

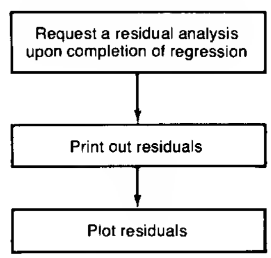
This program allows you to analyze the residuals from a regression problem in order to check the adequacy of the regression model. It may be used upon completion of any of the regression routines. The residuals may be printed and/or plotted.

The residual printout includes the observed values, predicted values, residuals, and standardized residuals. A final column shows which residuals are significantly large.

The residual plot allows you to plot the standardized residuals versus observation number or versus any of the variables in the model.

Residuals may be generated for subfiles which were not used in the determining the regression equation. This may be useful as a method of confirming the adequacy of the derived model.

Typical Program Flow



Special Considerations

Range of Standardized Residuals

The standardized residuals are plotted in a range from -5 to 5. If any standardized residuals are outside this range they will not be plotted, but a note showing the number of residuals off scale will be added to the plot.

Significance of Residuals

The last column in the residual table output shows which residuals are significantly large. In this column, two asterisks are printed for standardized residuals between two and three standard deviations away from zero. Similarly, three asterisks are printed for standardized residuals between three and four standard deviations away from zero, and four asterisks are printed for standardized residuals four or more standard deviations away from zero.

Distance Between X Tic Marks When Plotting

The first tic mark will coincide with the minimum X value. Every tic mark will be labeled. Hence, an uncluttered plot would contain no more than 10 tic marks.

Methods and Formulae

Suppose you wish to fit a regression model of the form:

$$Y = B0 + B1X1 + B2X2$$

where B0, B1, and B2 are the regression coefficients. We will call the nth predicted value for Y, y(n), the nth residual r(n), and the Jth observation of the Ith variable, D(I,J). We would then calculate the following:

1. Predicted Y: y(n) = b0 + b1*D(X1,n) + b2*D(X2,n), where b0, b1, and b2 are the predicted regression coefficients.
2. Residual: r(n) = D(Y,n) - y(n)
3. Standard error of residuals: Ser = (residual mean square) ↑ .5, where the residual mean square is calculated in the regression routine.
4. Standardized residual: SR(n) = r(n)/Ser

The residuals for a nonlinear regression are derived in a similar manner except that the nonlinear regression model is used to predict Y.

Example 1: Multiple Linear Regression

The data below will illustrate Multiple Linear Regression. The data consists of three variables, X1, X2 and the independent variable Y:

```
Are you going to use user defined transformation

or do Non-linear regression? (Y/N)
NO
Are you using an HPIB Printer?

YES

Enter select code, bus address (if 7,1 press CONT)?

*******************************************************************************
*                              DATA MANIPULATION                              *
*******************************************************************************

Enter DATA TYPE:
1                                                          Raw data
Mode number = ?

2                                                          Stored on mass storage
Is data stored on the program's scratch file (DATA)?

YES                                                        Previously stored on scratch data file.


                    EXAMPLE OF MULTIPLE LINEAR REGRESSION


Data file name: DATA

Data type is:    Raw data

Number of observations:      9
Number of variables:         6


Variable names:
   1. X1
   2. X2
   3. Y                                                    Note: X4, X5, and X6 are derived from X1
   4. X1^2                                                 and X2 by transformations.
   5. X2^2
   6. X1*X2

Subfiles:   NONE


SELECT ANY KEY
                                                           Select special function key labeled-LIST
Option number = ?
1                                                          List all the data

Enter method for listing data:
```

3 In tabular form

```
                    MULTIPLE LINEAR REGRESSION EXAMPLE

Data type is: Raw data

      Variable # 1   Variable # 2   Variable # 3   Variable # 4   Variable # 5
      (X1        )   (X2        )   (Y         )   (X1^2      )   (X2^2      )

OBS#
   1       7.80000        4.00000        0.00000       60.84000       16.00000
   2       7.80000        8.00000         .03100       60.84000       64.00000
   3       7.80000       12.00000         .47500       60.84000      144.00000
   4      39.00000        4.00000         .01600     1521.00000       16.00000
   5      39.00000        8.00000   8.000000E-03     1521.00000       64.00000
   6      39.00000       12.00000         .19000     1521.00000      144.00000
   7      78.00000        4.00000        0.00000     6084.00000       16.00000
   8      78.00000        8.00000         .03900     6084.00000       64.00000
   9      78.00000       12.00000        0.00000     6084.00000      144.00000


      Variable # 6
      (X1*X2     )

OBS#
   1      31.20000
   2      62.40000
   3      93.60000
   4     156.00000
   5     312.00000
   6     468.00000
   7     312.00000
   8     624.00000
   9     936.00000
```

For this data set only X1, X2 and Y need by typed in. When this is done, select the tranformation key on the template. To get X1 ↑ 2, choose option 1 allowing a = 1, b = 2, and c = 0. This creates a new variable X ↑ 2. The same is done to obtain X2 ↑ 2. To obtain X1*X2, choose option 10 allowing a = 1, b = 1, and c = 1. Once you have all these variables, store them by using the Store key on the template.

```
Option number = ?

0
SELECT ANY KEY
```

Exit from the List routine.

Select Special Function Key labeled-STATS

```
What statistic options are desired ?

1
VARIABLES =
?
```

Select just the mean, ci, variance, standard deviation, skewness, and kurtosis of all the data variables.

```
ALL
Confidence coefficient for confidence interval on the mean(e.g. 90,95,99%) = ?

95
*************************************************************************************
*                              SUMMARY STATISTICS                                   *
*                                 ON DATA SET:                                      *
*                        MULTIPLE LINEAR REGRESSION EXAMPLE                         *
*************************************************************************************
```

95% ci for means requested.

BASIC STATISTICS

VARIABLE NAME	# OF OBS.	# OF MISS	SUM	MEAN	VARIANCE	STD.DEV.
X1	9	0	374.40000	41.60000	927.81000	30.45997
X2	9	0	72.00000	8.00000	12.00000	3.46410
Y	9	0	.75900	.08433	.02506	.15832
X1^2	9	0	22997.52000	2555.28000	7403936.57637	2721.01756
X2^2	9	0	672.00000	74.66667	3136.00000	56.00000
X1*X2	9	0	2995.20000	332.80000	90043.20000	300.07199

VARIABLE NAME	COEFFICIENT OF VARIATION	STD. ERROR OF MEAN	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
X1	73.22109	10.15332	18.18009	65.01991
X2	43.30127	1.15470	5.33654	10.66346
Y	187.72946	.05277	-.03739	.20606
X1^2	106.48608	907.00585	463.15784	4647.40216
X2^2	75.00000	18.66667	31.60967	117.72366
X1*X2	90.16586	100.02400	102.08217	563.51783

VARIABLE	SKEWNESS	KURTOSIS
X1	.13506	-1.50000
X2	0.00000	-1.50000
Y	1.93769	2.29099
X1^2	.53922	-1.50000
X2^2	.29480	-1.50000
X1*X2	.88424	-.26334

```
What statistic options are desired ?

2
VARIABLES =
?
ALL
```

Request the correlation matrix of all the data variables.

```
*****************************************************************************
*                         SUMMARY STATISTICS                                *
*                            ON DATA SET:                                   *
*                  MULTIPLE LINEAR REGRESSION EXAMPLE                       *
*****************************************************************************
                          CORRELATION MATRIX
```

	X2	Y	X1^2	X2^2	X1*X2
X1	0.0000000	-.4209438	.9747877	0.0000000	.8120711
X2		.5916875	0.0000000	.9897433	.4802402
Y			-.3905355	.6250961	-.2314209
X1^2				0.0000000	.7915969
X2^2					.4753145

```
What statistic options are desired ?

3
VARIABLES =
?

ALL
```

Gives median, mode, percentiles, min, max, and range of all the data.

```
********************************************************************************
*                          SUMMARY STATISTICS                                  *
*                            ON DATA SET:                                      *
*                  MULTIPLE LINEAR REGRESSION EXAMPLE                          *
********************************************************************************
```

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
X1	78.00000	7.80000	70.20000	42.90000
X2	12.00000	4.00000	8.00000	8.00000
Y	.47500	0.00000	.47500	.23750
X1^2	6084.00000	60.84000	6023.16000	3072.42000
X2^2	144.00000	16.00000	128.00000	80.00000
X1*X2	936.00000	31.20000	904.80000	483.60000

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
X1	39.00000	7.80000	39.00000
X2	8.00000	4.00000	8.00000
Y	.01600	0.00000	.03100
X1^2	1521.00000	60.84000	1521.00000
X2^2	64.00000	16.00000	64.00000
X1*X2	312.00000	93.60000	312.00000

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
X1	40.56000	31.20000	31.20000
X2	8.00000	7.00000	4.00000
Y	.01880	.01575	.03100
X1^2	2141.56800	1155.96000	1460.16000
X2^2	70.40000	52.00000	48.00000
X1*X2	268.32000	257.40000	218.40000

```
Other percentiles?
NO
```

What statistic options are desired ? 0 SELECT ANY KEY	Note: All three sets of statistics could have selected original by answering ALL to option question. Exit Basic Statistics routine.
	Select special function key labeled-ADV STATS
Option number = ?	Remove BSDM medium. Insert regression medium.
1 Number of the dependent variable = ?	Multiple linear regression.
3 Which of the remaining variables should be included in the regression ?	Y = variable "Y"
ALL Is above information correct?	X_1, X_2, X ↑ 2, X2 ↑ 2, X1 and X2
YES	Displayed on CRT

```
*******************************************************************************
MULTIPLE LINEAR REGRESSION ON DATA SET:
                       MULTIPLE LINEAR REGRESSION EXAMPLE
*******************************************************************************

--where:   Dependent variable = (3)Y
           Independent variable(s) = (1)X1
                                     (2)X2
                                     (4)X1^2
                                     (5)X2^2
                                     (6)X1*X2
```

VARIABLE	N	MEAN	VARIANCE	STANDARD DEVIATION	COEFF. OF VARIATION
X1	9	41.60000	927.81000	30.45997	73.22109
X2	9	8.00000	12.00000	3.46410	43.30127
X1^2	9	2555.28000	7403936.57637	2721.01756	106.48608
X2^2	9	74.66667	3136.00000	56.00000	75.00000
X1*X2	9	332.80000	90043.20000	300.07199	90.16586
Y	9	.08433	.02506	.15832	187.72946

CORRELATION MATRIX

	X2	X1^2	X2^2	X1*X2	Y
X1	0.0000000	.9747877	0.0000000	.8120711	-.4209438
X2		0.0000000	.9897433	.4802402	.5916875
X1^2			0.0000000	.7915969	-.3905355
X2^2				.4753145	.6250961
X1*X2					-.2314209

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE
TOTAL	8	.20052		
REGRESSION	5	.17769	.03554	4.67
X1	1	.03553	.03553	4.67
X2	1	.07020	.07020	9.23
X1^2	1	.00158	.00158	.21
X2^2	1	.01531	.01531	2.01
X1*X2	1	.05507	.05507	7.24
RESIDUAL	3	.02283	.00761	

```
R-SQUARED =   .88615
STANDARD ERROR OF ESTIMATE = .0872327012721
```

From the AOV table we see that the additional sum of square for each variable produces a 'reasonable' F except X_4 and X_5.

VARIABLE	REGRESSION COEFFICIENTS STD. FORMAT	E-FORMAT	STANDARD ERROR REG. COEFFICIENT	T-VALUE
'CONSTANT'	-.00218	-.218154219795E-02	.25209	-.01
X1	.00247	.246964177292E-02	.00517	.48
X2	-.02576	-.257643442623E-01	.06364	-.40
X1^2	.00002	.231329291158E-04	.00005	.46
X2^2	.00547	.546875000000E-02	.00386	1.42
X1*X2	-.00083	-.833990121900E-03	.00031	-2.69

```
Confidence coefficient (e.g., 90,95,99) = ?
95
```

Note: All but the last T values are very small.
Not a very good model.

	COEFFICIENT	95 % CONFIDENCE INTERVAL LOWER LIMIT	UPPER LIMIT
'CONSTANT'	-.00218	-.72581	.72145
X1	.00247	-.01237	.01731
X2	-.02576	-.20845	.15692
X1^2	.00002	-.00012	.00017
X2^2	.00547	-.00560	.01653
X1*X2	-.00083	-.00172	.00006

```
Residual analysis and/or prediction ?
YES
Print out residuals?
YES
```

TABLE OF RESIDUALS

OBS#	OBSERVED Y	PREDICTED Y	RESIDUAL	STANDARDIZED RESIDUAL	SIGNIF
1	0.00000	-.02309	.02309	.26468	
2	.03100	.11033	-.07933	-.90944	
3	.47500	.41876	.05624	64476	
4	.01600	-.01634	.03234	.37073	
5	.00800	.01300	-.00500	- 05732	
6	.19000	.21734	-.02734	-.31342	
7	0.00000	.05543	-.05543	-.63541	
8	.03900	-.04533	08433	.96676	
9	0.00000	02890	-.02890	-.33135	

```
Durbin-Watson Statistic: 2.8245975174
```

For test for autocorrelation of residuals.

```
Residual plots?
YES
```

Residual Plots

```
Would you like to plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')
```

Press CONTINUE

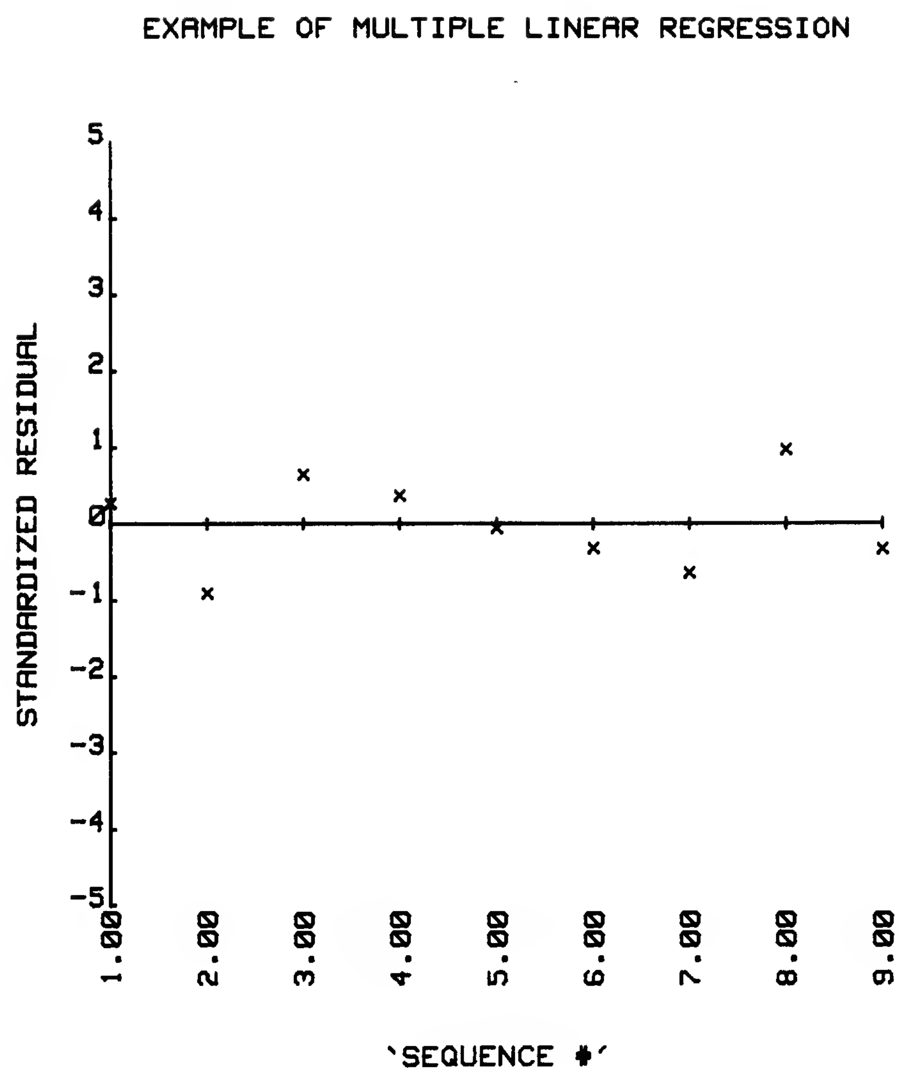
```
Plotter select code, Bus # = (defaults are 7,5)?
```

Press CONTINUE

```
Residual plot option no. = ?
1
```

Plot residuals vs time sequence.

```
For plotting, X-min = ?
1
For plotting, X-max = ?
9
Distance between X-ticks = ?
1
# of decimals for labelling X-axis (<=7) = ?
2
Number of pen color to be used ?
1
Is above information correct?
YES
```




```
Residual plots ?
0
Option number = ?
7
```

Exit from residual plots.

Return to BSDM.

Example 2: Stepwise Regression

The data shown below is the same as used in Multiple Linear Regression. Following the data are the results from the stepwise and backward selection procedures.

```
Are you going to use user defined transformation

or do Non-linear regression? (Y/N)

NO
Are you using an HPIB Printer?

YES
Printer select code, bus address = ?

Enter select code, bus address (if 7,) press CONT)?


*******************************************************************************************
*                              DATA MANIPULATION                                          *
*******************************************************************************************

Enter DATA TYPE:
1                                                              Raw data
Mode number = ?

2                                                              Stored on mass storage
Is data stored on the program's scratch file (DATA)?

YES                                                            Previously stored
                                                               Same as MLR example.


                    EXAMPLE OF STEPWISE LINEAR REGRESSION


Data file name: DATA

Data type is:   Raw data

Number of observations:     9
Number of variables:        6


Variable names:
    1. X1
    2. X2
    3. Y
    4. X1^2
    5. X2^2
    6. X1*X2

Subfiles:  NONE


SELECT ANY KEY
                                                               Select special function key labeled-LIST
Option number = ?

1                                                              List all the data.
Enter method for listing data:

3                                                              In tabular form.
```

EXAMPLE OF STEPWISE LINEAR REGRESSION

Data type is: Raw data

OBS#	Variable # 1 (X1)	Variable # 2 (X2)	Variable # 3 (Y)	Variable # 4 (X1^2)	Variable # 5 (X2^2)
1	7.80000	4.00000	0.00000	60.84000	16.00000
2	7.80000	8.00000	.03100	60.84000	64.00000
3	7.80000	12.00000	.47500	60.84000	144.00000
4	39.00000	4.00000	.01600	1521.00000	16.00000
5	39.00000	8.00000	8.000000E-03	1521.00000	64.00000
6	39.00000	12.00000	.19000	1521.00000	144.00000
7	78.00000	4.00000	0.00000	6084.00000	16.00000
8	78.00000	8.00000	.03900	6084.00000	64.00000
9	78.00000	12.00000	0.00000	6084.00000	144.00000

OBS#	Variable # 6 (X1*X2)
1	31.20000
2	62.40000
3	93.60000
4	156.00000
5	312.00000
6	468.00000
7	312.00000
8	624.00000
9	936.00000

This is the same data set that was used for multiple linear regression. Refer to that example for instructions on how to form X1 ↑ 2, X2 ↑ 2, X1*X2.

```
Option number = ?
0
SELECT ANY KEY
```

Exit the List routine.

Select special function key labeled-ADV STATS

Remove BSDM disc.

Insert Regression Medium.

```
Option number = ?
2
Procedure number = ?
1
Tolerance value (i.e. .01 .001) = ?
```

Stepwise regression

Choose the stepwise algorithm.

```
.01
F-value for inclusion = ?
```

Input tolerance value.

```
4
F-value for deletion = ?
```

F – to enter A F-value with 1 and n – k degrees of freedom where k = expected number of coefficients in mode.

f – to delete

```
4
Is above information correct?
YES
Number of dependent variable = ?
```

Note: We used F enter = F delete a common practice. Also, for n = 9 we probably should have used a much larger F. We definitely do not recommend small sample sizes except as examples.

```
3
Which remaining variables desired in regression?
ALL
```

Variable 3 = Y

With all others used as X_i.

```
Is above information correct?
YES
```

Information on CRT

```
*********************************************************************************
STEPWISE REGRESSION on DATA SET:
                              EXAMPLE OF STEPWISE LINEAR REGRESSION
*********************************************************************************
```

```
Dependent variable: (3)Y
Independent variable(s): (1)X1
                         (2)X2
                         (4)X1^2
                         (5)X2^2
                         (6)X1*X2
Tolerance = .01
F-value for inclusion = 4
F-value for deletion = 4
Method number = ?
2
```

The stepwise algorithm can enter or delete variables at a step. This example does not show any variables which are deleted.

CORRELATION MATRIX

	X1	X2	X1^2	X2^2	X1*X2	Y
X1	1.0000000	0.0000000	.9747877	0.0000000	.8120711	-.4209438
X2		1.0000000	0.0000000	.9892433	.4802402	.5916875
X1^2			1.0000000	0.0000000	.7915969	-.3905355
X2^2				1.0000000	.4753145	.6250961
X1*X2					1.0000000	-.2314209
Y						1.0000000

```
*********************************************************************************
STEP NUMBER 0
```

#--VARIABLE	F TO ENTER	PART CORR	TOL	F TO DELETE	REGRESSION COEFFICIENTS STD.FORMAT	E-FORMAT	STD ERROR
1.X1	1.51	.421	1.000				
2.X2	3.77	.592	1.000				
4.X1^2	1.26	.391	1.000				
5.X2^2	4.49	.625	1.000				
6.X1*X2	.40	.231	1.000				

Var. 5 has largest F-value and correlation, so it is the variable to enter the model.

```
*********************************************************************************
STEP NUMBER 1
VARIABLE'X2^2' ADDED
R-SQUARED =    39075
```

Analysis of Variance Table

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
TOTAL	8	.20052		
REGRESSION	1	.07835	.07835	4.49
RESIDUAL	7	.12217	.01745	

STANDARD ERROR = .132107402855

#--VARIABLE	F TO ENTER	PART CORR	TOL	F TO DELETE	REGRESSION COEFFICIENTS STD.FORMAT	E-FORMAT	STD ERROR
1.X1	2.46	.539	1.000				
2.X2	.37	.242	.020				
4.X1^2	2.00	.500	1.000				
5.X2^2				4.49	.00177	.176721538776E-02	.0008
6.X1*X2	8.72	.770	.774				

```
Constant = -.047619047619
```

Var 6 has the largest F-value and correlation, so it is the variable to enter the model.

```
*******************************************************************************
STEP NUMBER 2
VARIABLE'X1*X2' ADDED
R-SQUARED =   .75163

                     Analysis of Variance Table

SOURCE              DF         SUM OF SQUARES          MEAN SQUARE         F-VALUE
TOTAL                8                .20052
REGRESSION           2                .15072                 .07536           9.08
RESIDUAL             6                .04980                 .00830

STANDARD ERROR =  .0911067112552

                F TO   PART          F TO          REGRESSION COEFFICIENTS          STD
#--VARIABLE    ENTER   CORR     TOL DELETE    STD.FORMAT               E-FORMAT    ERROR
 1.X1           4.71   .696    .148
 2.X2            .45   .286    .020
 4.X1^2         4.53   .689    .190
 5.X2^2                              16.86       00268   .268474330203E-02       0007
 6.X1*X2                              8.72     -.00036 -.360245767615E-03       .0001

Constant =    .0037622915776
```

Var 1 has the largest F-value and correlation, so it is the variable to enter the model.

```
*******************************************************************************
STEP NUMBER 3
VARIABLE'X1' ADDED
R-SQUARED =   .87206

                     Analysis of Variance Table

SOURCE              DF         SUM OF SQUARES          MEAN SQUARE         F-VALUE
TOTAL                8                .20052
REGRESSION           3                .17486                 .05829          11 36
RESIDUAL             5                .02565                 .00513

STANDARD ERROR =  .0716294324428

                F TO   PART          F TO          REGRESSION COEFFICIENTS          STD
#--VARIABLE    ENTER   CORP     TOL DELETE    STD.FORMAT               E-FORMAT    ERPOR
 1.X1                                 4.71       00469   .468749152939E-02       0022
 2.X2            .20   .220    .020
 4.X1^2          .26   .248    .050
 5.X2^2                              25.74      .00396   .395611766121E-02      .0008
 6.X1*X2                             11.89     -.00086 - .85942300316-03        .0002

Constant = - .120040391928
```

None of the remaining variables have an F-value greater than F-To-Enter and none of the variables in the model have an F-value less than F-To-Delete, so the model is complete with X1, X2↑2, and X1*X2.

```
Tolerance value too small and/or F-values insufficient to proceed
Input 'K', delete '-K', or, enter 0 to end regression    . . .
0
```

No other terms added or removed.

```
Procedure number = ?
```

```
2
```

Choose the forward (stepwise) algorithm.

```
Tolerance value (i.e. .01..001) = ?
.01
```

Tolerance

```
F-value for inclusion = ?
4
```

F-To-Enter (perhaps too small)

```
Is above information correct?
```

```
YES
Number of dependent variable = ?
3
Which of the remaining variables should be used in the regression ?
ALL
Is above information correct?
YES
```

Note: No F to remove in FORWARD.

$Y = X_3$

All others potential.

```
********************************************************************************
FORWARD REGRESSION on DATA SET:
                  EXAMPLE OF STEPWISE LINEAR REGRESSION
********************************************************************************


Dependent variable: (3)Y
Independent variable(s): (1)X1
                         (2)X2
                         (4)X1^2
                         (5)X2^2
                         (6)X1*X2
Tolerance = .01
F-value for inclusion = 4
Method number = ?
2
```

The forward procedure will only add variables to the model and will stop when no variable has an F to enter larger than 4 (or whatever value you specify).

CORRELATION MATRIX

	X1	X2	X1^2	X2^2	X1*X2	Y
X1	1.0000000	0.0000000	.9747877	0.0000000	.8120711	-.4209438
X2		1.0000000	0.0000000	.9897433	.4802402	.5916875
X1^2			1.0000000	0.0000000	.7915969	-.3905355
X2^2				1.0000000	.4753145	.6250961
X1*X2					1.0000000	-.2314209
Y						1.0000000

```
********************************************************************************
STEP NUMBER 0
```

#--VARIABLE	F TO ENTER	PART CORR	TOL	F TO DELETE	REGRESSION COEFFICIENTS STD.FORMAT	E-FORMAT	STD ERROR
1.X1	1.51	.421	1.000				
2.X2	3.77	.592	1.000				
4.X1^2	1.26	.391	1.000				
5.X2^2	4.49	.625	1.000				
6.X1*X2	.40	.231	1.000				

The results for this portion of the example will be the same as the stepwise algorithm above.

```
********************************************************************************
STEP NUMBER 1
VARIABLE'X2^2' ADDED
R-SQUARED =  .39075
```

Analysis of Variance Table

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE
TOTAL	8	.20052		
REGRESSION	1	.07835	.07835	4.49
RESIDUAL	7	.12217	.01745	

STANDARD ERROR = .132107402855

#--VARIABLE	F TO ENTER	PART CORR	TOL	F TO DELETE	REGRESSION COEFFICIENTS STD.FORMAT	E-FORMAT	STD ERROR
1.X1	2.46	.539	1.000				
2.X2	.37	.242	.020				
4.X1^2	2.00	.500	1.000				
5.X2^2				4.49	.00177	.176721938776E-02	.0008
6.X1*X2	8.72	.770	.774				

Constant = -.047619047619

****
STEP NUMBER 2
VARIABLE'X1*X2' ADDED
R-SQUARED = .75163

Analysis of Variance Table

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE
TOTAL	8	.20052		
REGRESSION	2	.15072	.07536	9.08
RESIDUAL	6	.04980	.00830	

STANDARD ERROR = .0911067112552

#--VARIABLE	F TO ENTER	PART CORR	TOL	F TO DELETE	REGRESSION COEFFICIENTS STD.FORMAT	E-FORMAT	STD ERROR
1.X1	4.71	.696	.148				
2.X2	.45	.286	.020				
4.X1^2	4.53	.689	.190				
5.X2^2				16.86	00268	.268474330198E-02	.0007
6.X1*X2				8.72	-.00036	-.360245767605E-03	.0001

Constant = .0037622915776

****
STEP NUMBER 3
VARIABLE'X1' ADDED
R-SQUARED = .87206

Analysis of Variance Table

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE
TOTAL	8	20052		
REGRESSION	3	.17486	.05829	11.36
RESIDUAL	5	.02565	.00513	

STANDARD ERROR = .0716294324428

#--VARIABLE	F TO ENTER	PART CORR	TOL	F TO DELETE	REGRESSION COEFFICIENTS STD.FORMAT	E-FORMAT	STD ERROR
1.X1				4 71	.00469	.468749153034E-02	.0022
2.X2	.20	.220	.020				
4.X1^2	.26	.248	.050				
5.X2^2				25.76	.00396	.395611766121E-02	.0008
6.X1*X2				11.89	-.00086	-.859423200316E-03	.0002

Constant = - .120040391928

The results are the same as in stepwise regression.

Tolerance value too small and/or F-values insufficient to proceed.

```
Input 'K', delete '-K', or, enter 0 to end regression . . . .
0
Procedure number = ?

3
Tolerance value (i.e. .01..001) = ?

.01
F-value for deletion = ?

4
Is above information correct?
YES
Number of dependent variable = ?

3
Which remaining variables desired in regression ?

ALL
Is above information correct?
YES
```

Backward (stepwise) algorithm.

Only a F-To-Delete is required.

(Perhaps it should be bigger than 4 with n = 9.)

```
******************************************************************************
BACKWARD REGRESSION on DATA SET:
                    EXAMPLE OF STEPWISE LINEAR REGRESSION
******************************************************************************


Dependent variable: (3)Y
Independent variable(s): (1)X1
                         (2)X2
                         (4)X1^2
                         (5)X2^2
                         (6)X1*X2
Tolerance = .01
F-value for deletion = 4
Method number = ?
2
```

The backwards algorithm sets all the terms in the model and then deletes one at a time until no F to remove is less than the F we specify (Fdelete = 4).

CORRELATION MATRIX

	X1	X2	X1^2	X2^2	X1*X2	Y
X1	1.0000000	0.0000000	.9747877	0.0000000	.8120751	-.4202438
X2		1.0000000	0.0000000	.9897433	.4800409	.5916875
X1^2			1.0000000	0.0000000	.7915760	-.3905355
X2^2				1.0000000	.4754147	.6250961
X1*X2					1.0000000	-.0314209
Y						1.0000000

```
******************************************************************************
STEP NUMBER 0
R-SQUARED =   .88615
```

Analysis of Variance Table

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE
TOTAL	8	.20052		
REGRESSION	5	.17769	.03554	4.67
RESIDUAL	3	.02283	.00761	

STANDARD ERROR = .0872327012721

```
                  F TO   PART           F TO        REGRESSION COEFFICIENTS            STD
#---VARIABLE      ENTER  CORR     TOL. DELETE   STD.FORMAT            E-FORMAT        ERROR
 1.X1                                    .23       .00247   .246964177292E-02        .0052
 2.X2                                    .16      -.02576  -.257643442623E-01        .0636
 4.X1^2                                  .21       .00002   .231329291158E-04        .0001
 5.X2^2                                 2.01       .00547   .546875000000E-02        .0039
 6.X1*X2                                7.24      -.00083  -.833990121900E-03        .0003

Constant = -.00218154219794
```

Removes the variable with the smallest F to delete(x_2)

```
****************************************************************************************
STEP NUMBER 1
VARIABLE'X2' DELETED
R-SQUARED =  .87993

                         Analysis of Variance Table

SOURCE                DF         SUM OF SQUARES          MEAN SQUARE          F-VALUE
TOTAL                  8                 .20052
REGRESSION             4                 .17644               .04411             7.33
RESIDUAL               4                 .02408               .00602

STANDARD ERROR =  .0775817889132

                  F TO   PART           F TO        REGRESSION COEFFICIENTS            STD
#---VARIABLE      ENTER  CORR     TOL. DELETE   STD.FORMAT            E-FORMAT        ERROR
 1.X1                                    .34       .00267   .267310640025E-02        .0046
 2.X2              .16   .228    .020
 4.X1^2                                  .26       .00002   .231329291158E-04       0.0000
 5.X2^2                                21.96       .00396   .395611766121E-02        .0008
 6.X1*X2                               10.13      -.00086  -.859423200316E-03        .0003

Constant = -.0953530816668
```

Removes $X_4 = X1 \uparrow 2$ next.

```
****************************************************************************************
STEP NUMBER 2
VARIABLE'X1^2' DELETED
R-SQUARED =  .87206

                         Analysis of Variance Table

SOURCE                DF         SUM OF SQUARES          MEAN SQUARE          F-VALUE
TOTAL                  8                 .20052
REGRESSION             3                 .17486               .05829            11.36
RESIDUAL               5                 .02565               .00513

STANDARD ERROR =  .0716294324428
                  F TO   PART           F TO        REGRESSION COEFFICIENTS            STD
#---VARIABLE      ENTER  CORR     TOL. DELETE   STD.FORMAT            E-FORMAT        ERROR
 1.X1                                   4.71       .00469   .468749152939E-02        .0022
 2.X2              .20   .220    .020
 4.X1^2            .26   .248    .050
 5.X2^2                                25.76       .00396   .395611766121E-02        .0008
 6.X1*X2                               11.89      -.00086  -.859423200316E-03        .0002

Constant = -.120040391928
```

Results are the same as in stepwise regression.

```
Tolerance value too small and/or F-values insufficient to proceed.
```

But this may not be the case for some data sets.

```
Input 'K', delete '-K', or, enter 0 to end regression . . . .
0
Procedure number = ?
0
```

Exit Stepwise Regression.

```
Residual analysis and/or prediction?
NO
Option number = ?
7
```

Return to BSDM .

Example 3: Polynomial Regression

Bus Passenger Service Time

The time required to service boarding passengers at a bus stop was measured together with the actual number of passengers boarding. The service time was recorded from the moment that the bus stopped and the door opened until the last passenger boarded the bus. The objective is to determine a model for predicting passengers service time, given knowledge of the number boarding at a particular stop. Let Variable 1 = number boarding and Variable 2 = passenger service time. The following data was gathered during the month of May 1968 at twelve downtown locations in Louisville, Kentucky.

```
Are you going to use user defined transformation
or do Non-linear regression ? (Y/N)

NO
Are you using an HPIB Printer?

YES
Enter select code, bus address (if 7,1 press CONT)?



*******************************************************************************
*                              DATA MANIPULATION                              *
*******************************************************************************

Enter DATA TYPE:
1                                                        Raw data
Mode number = ?

2                                                        Mass storage
Is data stored on the program's scratch file 'DATA'?

YES                                                      Previously stored on 'Data File'



          BUS PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REGRESSION)


Data file name: DATA

Data type is:    Raw data

Number of observations:     31
Number of variables:         2


Variable names:
    1. NUMBER                                            X1 = number of passengers boarding a bus.
    2. TIME                                              X2 = Y = passenger service time in seconds.

Subfiles:  NONE


SELECT ANY KEY
                                                         Select special function key labeled-LIST
```

```
Option number = ?

1                                                    List all the data.
Enter method for listing data:

3                                                    In tabular form.
```

BUS PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REGRESSION)

Data type is: Raw data

OBS#	Variable # 1 (NUMBER)	Variable # 2 (TIME)
1	1.00000	1.40000
2	1.00000	2.80000
3	1.00000	3.00000
4	1.00000	1.80000
5	1.00000	2.00000
6	2.00000	4.70000
7	2.00000	8.00000
8	2.00000	3.00000
9	2.00000	2.50000
10	3.00000	5.20000
11	3.00000	6.20000
12	3.00000	9.40000
13	4.00000	11.70000
14	5.00000	7.50000
15	5.00000	11.90000
16	6.00000	13.60000
17	6.00000	12.40000
18	6.00000	11.60000
19	7.00000	14.70000
20	7.00000	13.50000
21	8.00000	12.00000
22	8.00000	14.10000
23	8.00000	26.00000
24	9.00000	19.00000
25	10.00000	21.20000
26	11.00000	22.90000
27	11.00000	22.60000
28	13.00000	25.20000
29	17.00000	33.50000
30	19.00000	33.70000
31	25.00000	54.20000

```
Option number = ?
0                                                    Exit List routine.
SELECT ANY KEY

                                                     Select special function key labeled-STATS
What statistic options are desired ?
1                                                    Gives the mean, ci, variance, standard, de-
VARIABLES =                                          viation, skewness, and kurtosis of all the
?                                                    data.
ALL
Confidence coefficient for confidence interval on the mean(e.g. 90,95,99%) = ?

95                                                   95%C.I.on means will be developed.
```

```
******************************************************************************
*                            SUMMARY STATISTICS                              *
*                               ON DATA SET:                                 *
*       BUS PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REGRESSION)        *
******************************************************************************


                              BASIC STATISTICS


 VARIABLE  # OF  # OF
  NAME     OBS.  MISS       SUM              MEAN          VARIANCE        STD.DEV.
NUMBER       31     0      207.00000         6.67742       33.22581         5.76418
TIME         31     0      431.30000        13.91290      139.39983        11.80677
------------------------------------------------------------------------------------


 VARIABLE    COEFFICIENT     STD. ERROR       95 % CONFIDENCE INTERVAL
  NAME       OF VARIATION     OF MEAN      LOWER LIMIT     UPPER LIMIT
NUMBER           86.32351       1.03528        4.56260         8.79223
TIME             84.86202       2.12056        9.58113        18.24468
------------------------------------------------------------------------------------


 VARIABLE          SKEWNESS          KURTOSIS

NUMBER              1.43125           1.90790
TIME                1.48977           2.55645
------------------------------------------------------------------------------------
What statistic options are desired ?
2
VARIABLES =
?
ALL
******************************************************************************
*                            SUMMARY STATISTICS                              *
*                               ON DATA SET:                                 *
*       BUS PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REGRESSION)        *
******************************************************************************
                             CORRELATION MATRIX


                   TIME
NUMBER          .9743533


------------------------------------------------------------------------------------
What statistic options are desired ?
3
VARIABLES =
?
ALL
******************************************************************************
*                            SUMMARY STATISTICS                              *
*                               ON DATA SET:                                 *
*       BUS PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REGRESSION)        *
******************************************************************************


                              ORDER STATISTICS

VARIABLE           MAXIMUM           MINIMUM            RANGE          MIDRANGE
NUMBER           25.00000           1.00000         24.00000         13.00000
TIME             54.20000           1.40000         52.80000         27.80000
------------------------------------------------------------------------------------
```

Gives the correlation matrix of all the data.

Highly correlated in a linear fashion.

Gives median, mode, percentiles, min, max, and range of all the data.

```
                                                    TUKEY'S HINGES
VARIABLE                      MEDIAN         25-th %-ile          75-th %-ile
NUMBER                       6.00000            2.00000              8.00000
TIME                        11.90000            4.70000             19.00000
--------------------------------------------------------------------------------
                                        TUKEY'S MIDDLEMEANS
VARIABLE                     MIDMEAN            TRIMEAN            MIDSPREAD
NUMBER                       5.41176            5.50000              6.00000
TIME                        11.57059           11.87500             14.30000
--------------------------------------------------------------------------------
Other percentiles?
NO
--------------------------------------------------------------------------------
What statistic options are desired ?
0
SELECT ANY KEY
```

Exit Basic Statistics.
Select special function key labeled-ADV STATS
Remove BSDM disc.
Insert regression medium.

```
Option number = ?
3
```

Polynomial regression selected.

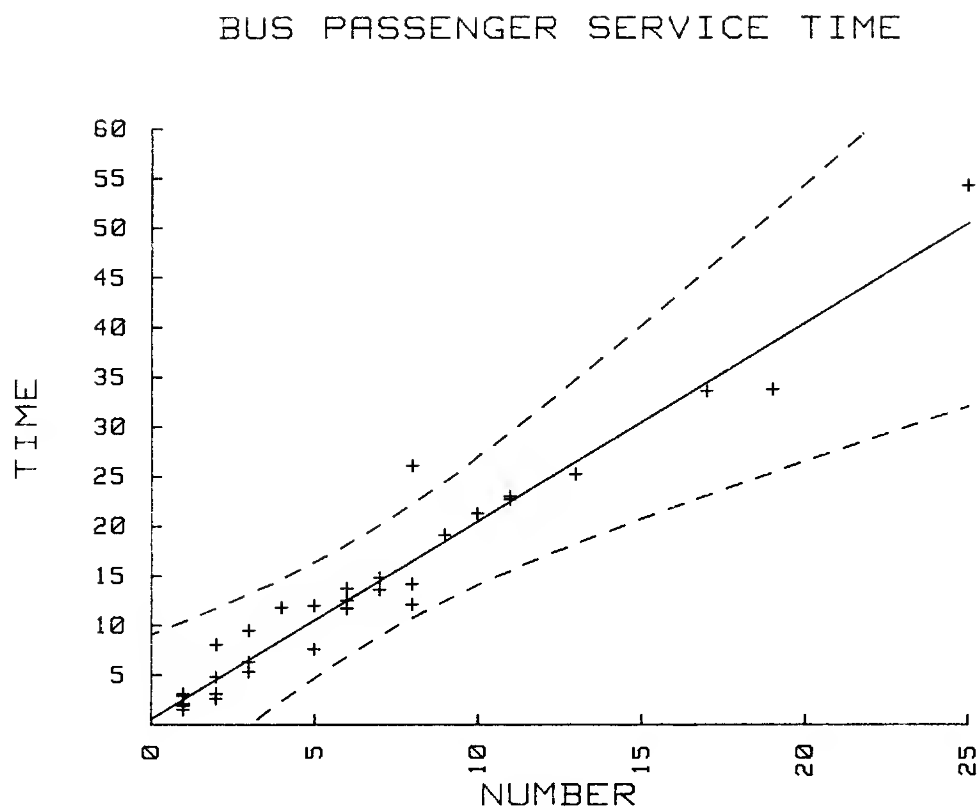
```
Number of the dependent variable = ?
2
Number of the independent variable = ?
1
*********************************************************************************
POLYNOMIAL REGRESSION ON DATA SET:
          BUS PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REGRESSION)
*********************************************************************************

--where:  Dependent variable = (2)TIME
          Independent variable = (1)NUMBER


Is a plot of the regression desired?
YES
Plot on CRT?
NO
```

Plot on an external plotter

```
Plotter identifier string (press CONT if 'HPGL') ?
Plotter select code, Bus # = (defaults are 7,5) ?
X-min = ?
0
X-max = ?
25
Y-min = ?
0
Y-max = ?
60
```

Plotting limits specified.

```
Y-axis crosses X-axis at X = ?
0
X-axis crosses Y-axis at Y = ?
0
Distance between X-ticks = ?
5
Distance between Y-ticks = ?
5
# of decimals for labelling X-axis (<=7) = ?
0
# of decimals for labelling Y-axis = ?
0
Number of pen color to be used ?
1
Is above information correct?
YES
Beep will sound when plot is done, then press CONTINUE
```


BUS PASSENGER SERVICE TIME
TIME
60
55
50
45
40
35
30
25
20
15
10
5
0
5
10
15
20
25
NUMBER

```
Maximum degree of regression(<=10) = ?
1
```

We specified maximum degree at 1 although we could have chosen a value slightly higher than desired level.

```
                                                              STANDARD      COEFF. OF
VARIABLE        N           MEAN            VARIANCE         DEVIATION      VARIATION
NUMBER         31        6.67742            33.22581           5.76418       86.32351
TIME           31       13.91290           139.39983          11.80677       84.86202

CORRELATION =    .97435


Degree of regression = ?
1
SELECTED DEGREE OF REGRESSION =  1
R-SQUARED =   .94936
STANDARD ERROR OF ESTIMATE = 2.70221890497
```

Specify the actual degree of interest.

```
                          ANALYSIS OF VARIANCE TABLE

SOURCE              DF          SUM OF SQUARES          MEAN SQUARE          F-VALUE

TOTAL               30             4181.99484
REGRESSION           1             3970.23722           3970.23722            543.72
 X^1                 1             3970.23722           3970.23722            543.72
RESIDUAL            29              211.75762              7.30199

                       REGRESSION COEFFICIENTS           STANDARD ERROR
VARIABLE           STD. FORMAT          E-FORMAT        REG. COEFFICIENT      T-VALUE
'CONSTANT'             .58633   .586330096900E+00             .74979            .78
 X^1                  1.99577   .199576699031E+01             .08559          23.32
Confidence coefficient (e.g., 90,95,99) = ?
95
```

$\hat{y} = .586 + 2.00X$ about two seconds per passenger to board a bus.

```
                                       95 % CONFIDENCE INTERVAL
                     COEFFICIENT         LOWER LIMIT         UPPER LIMIT
'CONSTANT'                .58633            -.94752             2.12018
 X^1                     1.99577            1.82068             2.17086
```

May not need an intercept term.

```
Plot regression curve on present graph ?
YES
Plot confidence interval of regression line also ?
YES
Confidence coefficient (e.g., 90, 95, 99)= ?
95
Same pen color ?
YES
Change degree of regression ?
NO
Residual analysis and/or prediction ?
YES
Print out residuals?
YES
```

TABLE OF RESIDUALS

OBS#	OBSERVED Y	PREDICTED Y	RESIDUAL	STANDARDIZED RESIDUAL	SIGNIF
1	1.40000	2.58210	-1.18210	-.43745	
2	2.80000	2.58210	.21790	.08064	
3	3.00000	2.58210	.41790	.15465	
4	1.80000	2.58210	-.78210	-.28943	
5	2.00000	2.58210	-.58210	-.21541	
6	4.70000	4.57786	.12214	.04520	
7	8.00000	4.57786	3.42214	1.26642	
8	3.00000	4.57786	-1.57786	-.58391	
9	2.50000	4.57786	-2.07786	-.76895	
10	5.20000	6.57363	-1.37363	-.50833	
11	6.20000	6.57363	-.37363	-.13827	
12	9.40000	6.57363	2.82637	1.04594	
13	11.70000	8.56940	3.13060	1.15853	
14	7.50000	10.56517	-3.06517	-1.13431	
15	11.90000	10.56517	1.33483	.49398	
16	13.60000	12.56093	1.03907	.38452	
17	12.40000	12.56093	-.16093	-.05956	
18	11.60000	12.56093	-.96093	-.35561	
19	14.70000	14.55670	.14330	.05303	
20	13.50000	14.55670	-1.05670	-.39105	
21	12.00000	16.55247	-4.55247	-1.68471	
22	14.10000	16.55247	-2.45247	-.90757	
23	26.00000	16.55247	9.44753	3.49621	***
24	19.00000	18.54823	.45177	.16718	
25	21.20000	20.54400	.65600	.24276	
26	22.90000	22.53977	.36023	.13331	
27	22.60000	22.53977	.06023	.02229	
28	25.20000	26.53130	-1.33130	-.49267	
29	33.50000	34.51437	-1.01437	-.37538	
30	33.70000	38.50590	-4.80590	-1.77850	
31	54.20000	50.48050	3.71950	1.37646	

```
Durbin-Watson Statistic: 2.09200089648
```

Note that one observation (#23) seems to have a very large standardized residual.

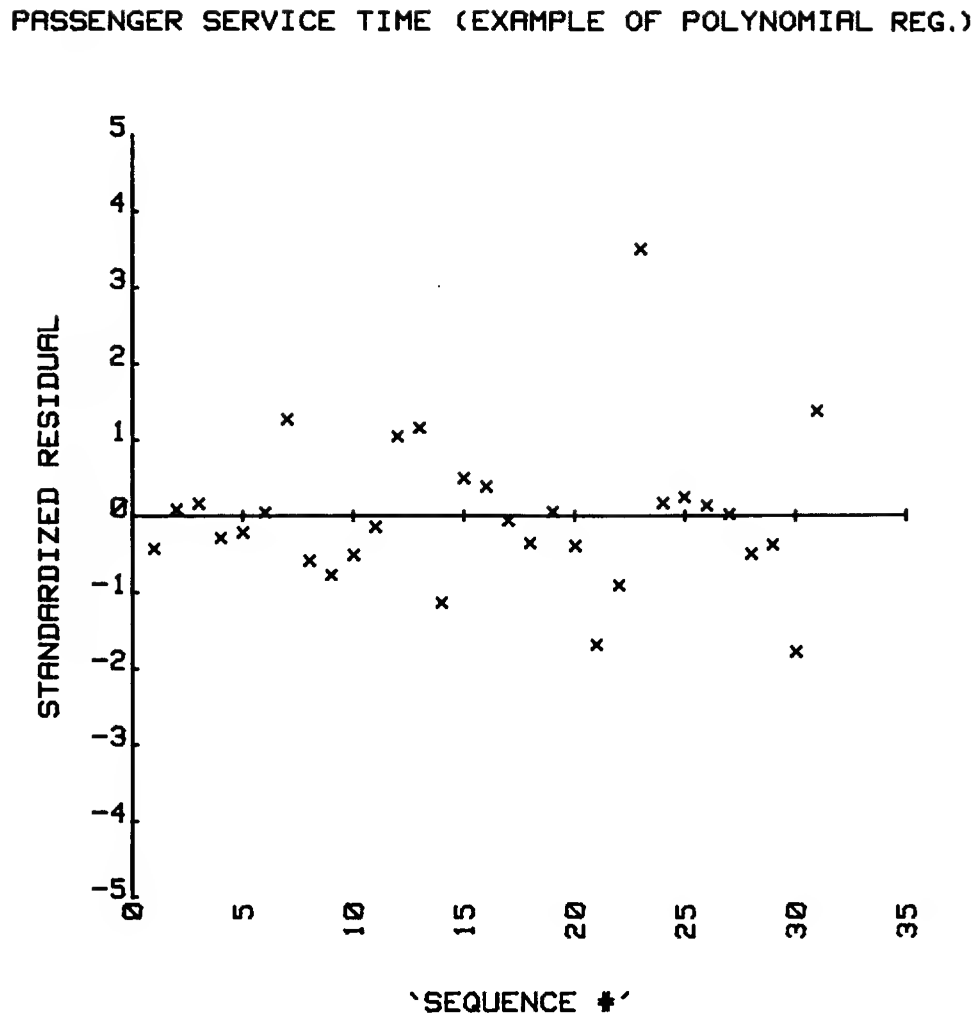
```
Residual plots?
YES
Plot on CRT?
NO
Plotter identifier string (press CONT if 'HPGL' ?
Plotter select code, Bus # = (defaults are 7,5)
```

Residual plots

An external plotter is used.

```
Residual plot option no. = ?
1
For plotting, X-min = ?
0
For plotting, X-max = ?
35
Distance between X-ticks = ?
5
# of decimals for labelling X-axis (<=7) = ?
0
Number of pen color to be used ?
1
Is above information correct?
YES
```

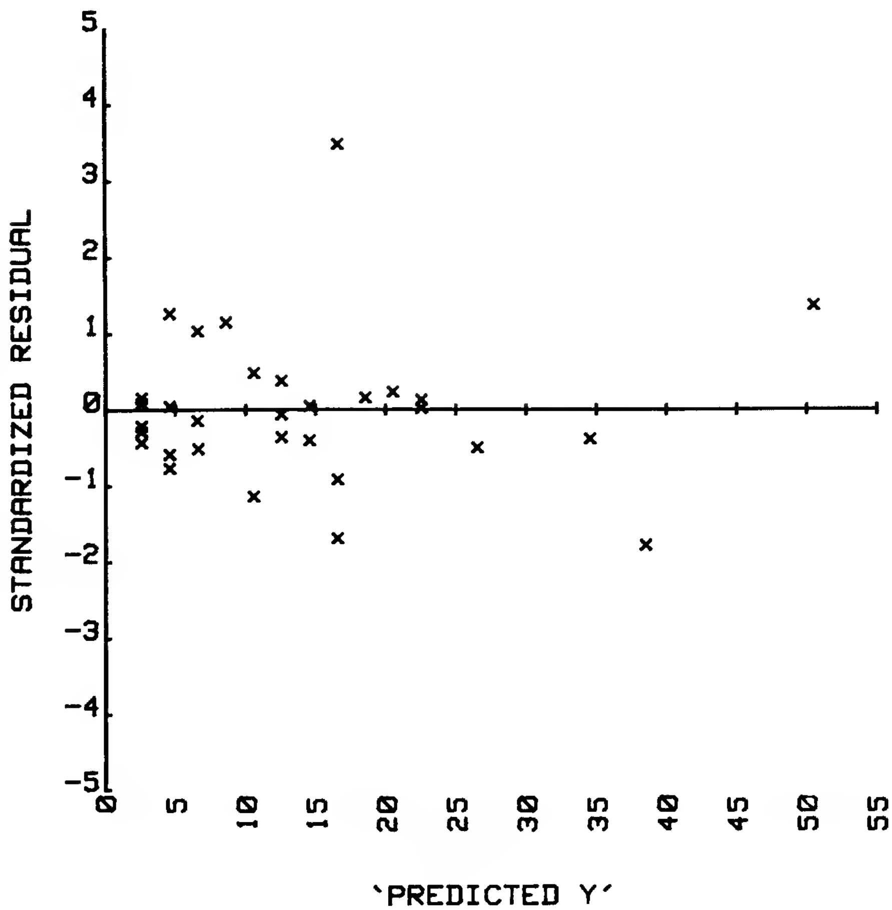
Plot residuals vs time sequence.




```
Residual plots ?
YES
Plotter identifier string (press CONT if 'HPGL') ?
Plotter select code, bus # (defaults are 7,5) ?
Residual plot option no. = ?
2
For plotting, X-min = ?
0
For plotting, X-max = ?
55
Distance between X-ticks = ?
5
# of decimals for labelling X-axis (<=7) = ?
0
Number of pen color to be used ?
1
Is above information correct?
YES
```

Plot residuals vs predicted Y values.

PASSENGER SERVICE TIME (EXAMPLE OF POLYNOMIAL REG.)



Residual plots ?
NO
Option number = ?
7

Return to BSDM .

Example 4: Nonlinear Regression

Twenty-five samples of human urine were obtained to determine if a nonlinear model could be developed relating Y = blood concentration of urine (micrograms/1000 cc) to X = time in hours.

The data were entered from the keyboard.

A "three-exponential" model was tried:

Yhat = B0*exp(− B1*X) + B2*exp(− B3*X) + B4*exp(− B5*X)
and 0.00001 was used as the convergence coefficient.

Notes:

1. The initial estimates were chosen after some experimentation although the only effect that they have is in the speed of convergence.
2. Every iteration was printed. It is not necessary to have this done.
3. The residuals for the smallest time are larger than for T or X near 60 or above. Of course, the largest Y's are associated with the smallest X.

```
Are you going to use user defined transformation
or do Non-linear regression ? (Y/N)

NO
Are you using an HPIB Printer?

YES
Enter select code, bus address (if 7,1 press CONT) ?
```

We have already prepared the file with the function and derivative.

```
*******************************************************************************
*                              DATA MANIPULATION                               *
*******************************************************************************

Enter DATA TYPE:
1
Mode number = ?

2
Is data stored on the program's scratch file (DATA)?

YES
```

Raw data data type required

From mass storage

Data stored in program's storage medium from previous run.

```
   EXAMPLE 1-URINE/BLOOD CONCENTRATION

Data file name: DATA

Data type is:    Raw data

Number of observations:     25
Number of variables:         2
```

```
Variable names:
   1  TIME(HR)
   2. BLD.CONT

Subfiles:  NONE


SELECT ANY KEY
```

Select special function key labeled-LIST

```
Option number = ?

1
```

List all the data.

```
Enter method for listing data:
3
```

In tabular form.

```
   EXAMPLE 1-URINE/BLOOD CONCENTRATION

Data type is: Raw data
```

OBS#	Variable # 1 (TIME(HR))	Variable # 2 (BLD.CONT)
1	4.25000	1165.70000
2	7.50000	851.00000
3	10.80000	523.00000
4	12.00000	365.00000
5	16.00000	294.00000
6	23.80000	170.00000
7	27.80000	60.00000
8	35.30000	81.00000
9	38.30000	20.00000
10	45.30000	45.00000
11	51.30000	27.00000
12	54.20000	37.00000
13	59.80000	31.00000
14	64.25000	26.00000
15	69.50000	36.00000
16	78.20000	18.00000
17	90.20000	10.00000
18	100.00000	8.20000
19	105.00000	13.40000
20	108.00000	17.40000
21	114.00000	8.00000
22	120.00000	4.00000
23	130.00000	6.70000
24	142.00000	6.70000
25	154.00000	5.80000

```
Option number = ?
0
SELECT ANY KEY
```

Exit the List routine.

Select special function key labeled-ADV STAT
Remove BSDM medium.
Insert the regression medium.

```
Option number = ?

4
Number of the dependent variable = ?
```

Select non-linear regression.

```
2
How many independent variables will be in the model?
```

Specify blood content as Y.

```
1
Independent variable numbers (separated by commas) ?
?
```

One independent variable.

Specify time in hours as X.

```
1
Is above information correct?

YES
************************************************************************************
NON-LINEAR REGRESSION ON DATA SET:
        URINE/BLOOD CONCENTRATION (EXAMPLE 1 OF NON-LINEAR REGRESSION)
************************************************************************************

---where:  Dependent variable = (2)BLD.CONT
           Independent variable(s) = (1)TIME(HR)

# of parameters in the model((=20)  ?

6
Is a plot of the non-linear regression desired

YES
Plot on CRT
```

Request plot

```
NO
Plotter identifier string (press CONT if 'HPGL') ?
Plotter select code,Bus# =(defaults are 7,5) ?
```

But not on CRT.

On plotter with select code = 7 and bus code = 5.

```
Is a quick plot desired ?

NO
X-min = ?
```

No quick plot. We will specify our limits.

```
4
X-max = ?

160
Y-min = ?

3
Y-max = ?

1170
Y-axis crosses X-axis at X = ?

4
X-axis crosses Y-axis at Y = ?

3
Distance between X-ticks = ?

16
Distance between Y-tic = ?
```

Xmin = 4
Xmax = 160
Ymin = 3
Ymax = 1170

Xtic inverval = 16
Ytic interval = 120
With no decimal points for labelling.

```
120
 # of decimals for labelling X-axis/(<7) ?

0
 # of decimals for labelling Y-axis ?

0
Number of pen color to be used ?
1

Is above information correct ?

YES


Beep will sound when plot done, then press CONTINUE
File name where subroutines are stored ?
FCNDER:INTERNAL
Is function medium placed in device?
YES
Is program medium placed in device?
YES

Enter convergence coefficient (e.g. 0.005,.001)

.00001
Initial estimate for parameter # 1
?

1202.336
Initial estimate for parameter # 2
?

.1083
Initial estimate for parameter # 3
?

400.3367
Initial estimate for parameter # 4
?

.1083
Initial estimate for parameter # 5
?

31.4619
Initial estimate for parameter # 6
?

.006716
Is the above information correct?

YES
*************************************************************************************
```

Convergence criteria on changes in all coefficients. Note .00001 is pretty restrictive.

Initial estimates input at this point.

```
Delta(Convergence criteria)=  .00001

THE INITIAL VALUES OF PARAMETERS ARE
         PARAMETER 1 = 1202.336
         PARAMETER 2 = .1083
         PARAMETER 3 = 400.3367
         PARAMETER 4 = .1083
         PARAMETER 5 = 31.4619
         PARAMETER 6 = .006716


Would you like to print out every iteration on hard copy option printer
YES
```

Not a good idea if many iterations are expected.

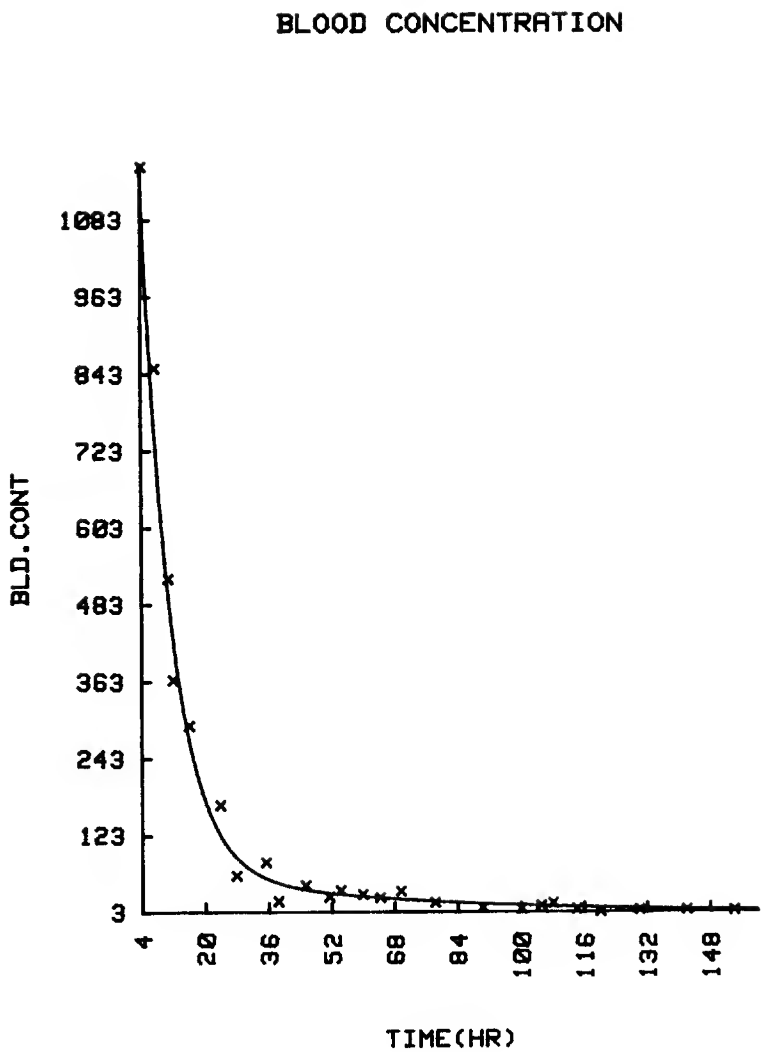
```
Calcs. may be lengthy. A beep will sound when done. Press 'S' key to ABORT!
ITERATION            ESTIMATED PARAMETER VALUES             S.S.RESIDUALS


Calculations may be quite time consuming. A beep will sound when completed.
    0     1202.33600        .10830      400.33670       .10830
          31.46190        .00672                           42560.6966977
    1     1379.00339        .12722      577.00409       .16513
          76.16355        .02198                           19113.9722052
    2     1392.99446        .13353      600.25867       .13849
          71.83127        .01538                           17230.2902581
    3     1395 63956        .14371      603.73447       .12102
          76.36979        .01725                           17131.1484877
    4     1397.91748        .14022      603.92050       .13013
          76.09567        .01722                           17001.3543193
    5     1398.50753        .13844      604.30809       .13435
          75.57048        .01714                           16990.4904512
    6     1398.59945        .13768      604.39161       .13606
          75.28321        .01709                           16989.3523974
    7     1398.59229        .13736      604.38569       .13672
          75.15983        .01707                           16989.1984944
    8     1398.58144        .13724      604.37522       .13698
          75.10969        .01706                           16989.1746607
    9     1398.57589        .13719      604.36975       .13708
          75.08959        .01706                           16989.1708856
   10     1398.57350        .13717      604.36737       .13713
          75.08157        .01706                           16989.1702861
   11     1398.57252        .13716      604.36639       .13714
          75.07838        .01706                           16989.1701889
   12     1398.57212        .13716      604.36599       .13715
          75.07711        .01706                           16989.1701720
   13     1398.57196        .13715      604.36583       .13715
          75.07660        .01706                           16989.1701669
 DONE!!!!
```

Note: Estimated values for six coefficients followed by sum of squared residuals.

```
*************************************************************************************
THE ESTIMATED PARAMETER VALUES AFTER 13 ITERATIONS ARE :
         PARAMETER  1=     1398.5719009 ( 1.3985719009E+03)
         PARAMETER  2=        .1371535 ( 1.3715347965E-01)
         PARAMETER  3=      604.3657684 ( 6.0436576836E+02)
         PARAMETER  4=        .1371525 ( 1.3715246328E-01)
         PARAMETER  5=       75.0763988 ( 7.5076398794E+01)
         PARAMETER  6=        .0170560 ( 1.7055987670E-02)
*************************************************************************************
          THE INITIAL VALUE OF SUM OF SQUARED RESIDUALS = 42560.6966977
          AFTER 13 ITERATIONS THE SUM OF SQUARED RESIDUALS= 16989.1701669
          APPROXIMATE STANDARD ERROR FROM SQUARED RESIDUALS= 29.9026228093
Plot regression curve on present GRAPH ?
YES
```

Plot curve to see how good the fit is.

```
Same pen color ?
YES
```




```
Like to change initial estimates and/or function ?

NO                                                            We are satisfied.
Are confidence intervals on parameters desired ?

YES                                                           Request confidence intervals.
Confidence coefficient for confidence interval on parameters(e.g. 90 95 99).

95
```

```
*******************************************************************************
              APPROXIMATE 95 % CONFIDENCE INTERVALS ON PARAMETERS
```

PARAMETER	ONE-AT-A TIME C.I.		SIMULTANEOUS C.I.	
	LOWER LIMIT	UPPER LIMIT	LOWER LIMIT	UPPER LIMIT
1	790.3196	2006.8242	244.5233	2552.6205
2	.0762	.1981	.0215	.2528
3	-3.8858	1212.6173	-549.6815	1758.4130
4	-.0039	.2782	-.1304	.4047
5	-33.5338	183.6866	-130.9917	281.1445
6	-.0073	.0414	-.0292	.0633

```
*******************************************************************************
Residual analysis and/or prediction?

YES
Print out residuals?
```

Study size and form of residuals.

```
YES
```

TABLE OF RESIDUALS

OBS#	OBSERVED Y	PREDICTED Y	RESIDUAL	STANDARDIZED RESIDUAL	SIGNIF
1	1165.70000	1188.01983	-22.31983	-.74642	
2	851.00000	782.09103	68.90897	2.30445	**
3	523.00000	517.81851	5.18149	.17328	
4	365.00000	447.44910	-82.44910	-2.75725	**
5	294.00000	280.31284	13.68716	.45772	
6	170.00000	126.59139	43.40861	1.45167	
7	60.00000	90.96313	-30.96313	-1.03547	
8	81.00000	56.93086	24.06914	.80492	
9	20.00000	49.54572	-29 54572	-.98806	
10	45.00000	38.68203	6.31797	.21128	
11	27.00000	33.05932	-6.05932	- 20264	
12	37.00000	30.97073	6.02927	.20163	
13	31.00000	27.62273	3.37727	.11294	
14	26.00000	25.39305	.60695	.02030	
15	36.00000	23.09053	12.90947	.43172	
16	18.00000	19.82514	-1.82514	-.06104	
17	10.00000	16.12840	-6.12840	-.20495	
18	8.20000	13.64085	-5.44085	-.18195	
19	13.40000	12.52486	.87514	.02927	
20	17.40000	11.89978	5.50022	.18394	
21	8.00000	10.74190	-2.74190	- 09169	
22	4.00000	9.69684	-5.69684	-.19051	
23	6.70000	8.17622	-1.47622	-.04937	
24	6.70000	6.66290	.03710	.00124	
25	5.80000	5.42969	.37031	01238	

These two have fairly large residuals.

```
Durbin-Watson Statistic: 2.57626883803

Residual plots?

YES
Plot on CRT?
```

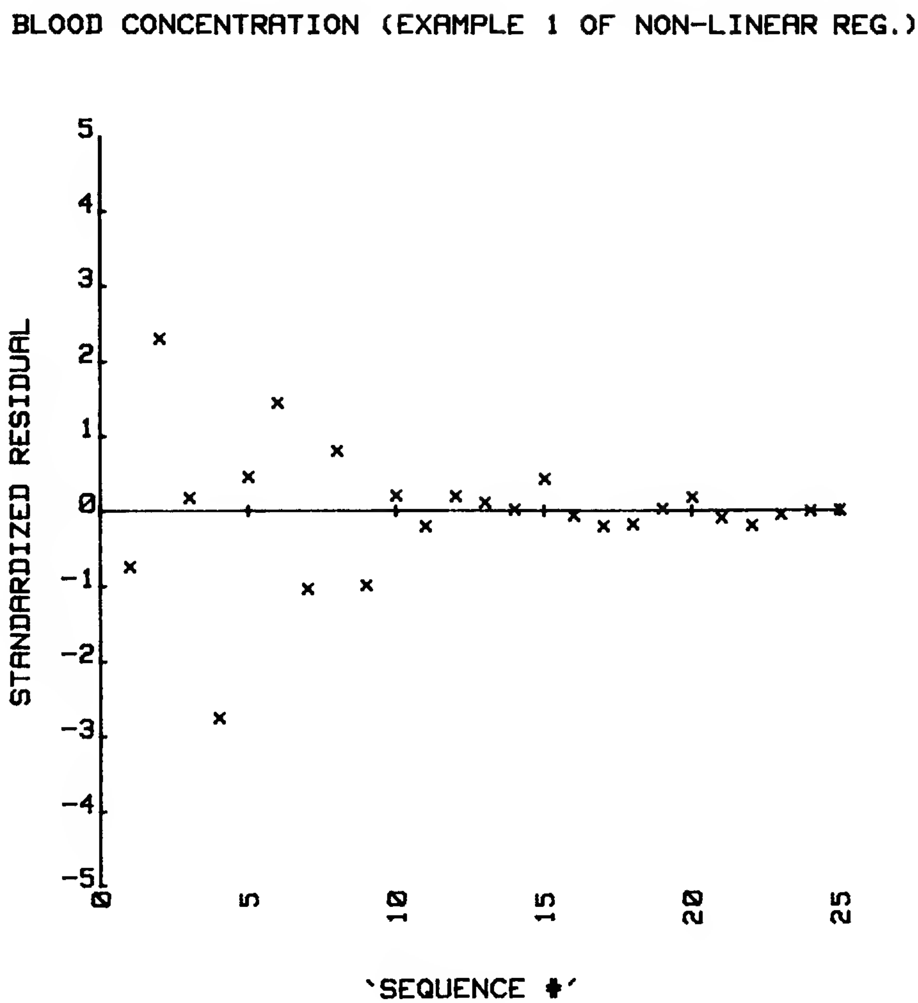
Residual plots yes

```
NO
```

On external plotter

```
Plotter identifier string (CONT if 'HPGL') ?
Plotter select code, Bus # = (defaults are 7,5) ?
```

```
Residual plot option no. = ?

1                                                    Plot residuals vs time/sequence number.
For plotting, X-min = ?

0
For plotting, X-max = ?

25
Distance between X-ticks = ?

5
# of decimals for labelling X-axis (<=7) = ?

0
Number of pen color to be used ?
1
Is above information correct?

YES
```




```
Residual Plots ?
NO                                                   Exit residual routine.

Option number = ?

7                                                    Return to BSDM
```

Example 5: Nonlinear Regression

An experiment was conducted to determine the relationship between Y = elevation (in centimeters) and X = distance from the summit of a hill.

Thirty-four observations were entered from a mass storage device.

After viewing the X-Y scatter plot, it appeared that it would be necessary to piece the model together. Hence, the following model suggested itself:

Yhat = polynomial model of degree 2 if X≤65.
= simple linear model if 65<X≤125.
= polynomial model of degree 2 if X>125.

i.e., the model can be written as

Y hat = A0 + A1*X + A2*X ↑ 2 if X≤65.
= B0 + B1*X if 65<X≤125.
= C0 + C1*X + C2*X ↑ 2 if X>125.

or for the program's purpose:

```
F= (P(1)+P(2)*X(1)+P(3)*X(1)↑2)*(X(1)<=65)+
   (P(4)+P(5)*X(1))*((X(1)>65)AND(X(1)<=125))+
   (P(6)+P(7)*X(1)+P(8)*X(1)↑2)*(X(1)>125)
```

Therefore, we have eight unknown parameters in the model to be estimated. 0.00001 was used as the convergence coefficient.

The initial values were obtained by interpolating values on the scatter plot. The chosen values are:

Initial Values:

A0 = 1000	B0 = 1200	C0 = 1826
A1 = −1.0	B1 = −5.8	C1 = −16.0
A2 = −.2		C2 = .046

After five iterations, the estimated coefficients give a Sum of Squares residual of about 295 and a very good fit as we can observe from the plot of the data and the estimated equation. Also, the residual analysis seems to suggest that the fit is quite good.

```
Are you going to use user defined transformation
or do Non-linear regression ? (Y/N)

NO
Are you using an HPIB Printer?

YES
Enter select code, bus address (if 7,1 press CONT) ?
```

Other printer selected.

```
*******************************************************************************
*                           DATA MANIPULATION                                 *
*******************************************************************************

Enter DATA TYPE:
```

```
1
Mode number = ?
```

Raw data (data type required)

```
2
Is data stored on the program's scratch file (DATA)?
```

From mass storage

```
NO
Data file name = ?
```

Data stored on a different medium so it must be retrieved.

```
LANDSCAPE: INTERNAL
Was data stored by the BS&DM system ?

YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in device?
YES
PROGRAM NOW STORING DATA ON SCRATCH DATA FILE AND BACKUP FILE



     LANDSCAPE SEGMENTS DELINEATION

Data file name: LND120:F8,1

Data type is:    Raw data

Number of observations:      34
Number of variables:          2


Variable names:
    1. DISTANCE
    2  ELEVATION

Subfile name       beginning observation--number of observations
 1  TOP                             1                        15
 2. BOTTOM                         16                        19

SELECT ANY KEY
```

Select special function key labeled-LIST

```
Option number = ?
```

```
1
```

List all the data

```
Enter method for listing data:
3
```

In tabular form

LANDSCAPE SEGMENTS DELINEATION

Data type is: Raw data

OBS#	Variable # 1 (DISTANCE)	Variable # 2 (ELEVATION)
1	0.00000	1000.00000
2	5.00000	992.40000
3	10.00000	985.40000
4	15.00000	973.30000
5	20.00000	963.10000
6	25.00000	952.90000
7	30.00000	939.60000
8	35.00000	929.40000
9	40.00000	912.90000
10	45.00000	894.50000
11	50.00000	881.80000
12	55.00000	864.00000
13	60.00000	832.90000
14	65.00000	808.80000
15	70.00000	779.00000
16	75.00000	757.40000
17	80.00000	727.60000
18	85.00000	691.40000
19	90.00000	664.10000
20	95.00000	633.00000
21	100.00000	605.70000
22	105.00000	577.10000
23	110.00000	549.80000
24	115.00000	518.00000
25	120.00000	495.10000
26	125.00000	468.40000
27	130.00000	446.20000
28	135.00000	421.40000
29	140.00000	403.00000
30	145.00000	390.90000
31	150.00000	369.30000
32	155.00000	356.60000
33	160.00000	347.70000
34	165.00000	340.10000

```
Option number = ?
0                                              Exit List routine.
SELECT ANY KEY
                                               Remove Basic Statistics
                                               Go to Regression program medium.

Option number = ?
4                                              Non-linear regression
Subfile #(enter 0 to ignore subfiles) = ?

0
Number of the dependent variable = ?

2
How many independent variables will be in the model?

1
Independent variable numbers (separated by commas) =
?
1
```

```
Is above information correct?

YES
********************************************************************************
NON-LINEAR REGRESSION ON DATA SET·
                              LANDSCAPE SEGMENTS DELINEATION
********************************************************************************

--where:  Dependent variable = (2)ELEVATION
          Independent variable(s) = (1)DISTANCE

# of parameters in the model((=20)  ?

8
Is a plot of the non-linear regression desired

YES
Plot on CRT

NO
Plotter identifier string (CONT if 'HPGL')?
Plotter select code,Bus# =(defaults are 7,5) ?
```

Plot on EXTERNAL plotter

```
Is a quick plot desired ?

NO
```

No quick plot. We specify our limits.

```
X-min = ?

0
X-max = ?

165
Y-min = ?

340
Y-max = ?

1000
Y-axis crosses X-axis at X = ?

0
X-axis crosses Y-axis at Y = ?

340
Distance between X-ticks = ?

33
Distance between Y-tic = ?

100
 # of decimals for labelling X-axis((=7) = ?

0
 # of decimals for labelling Y-axis = ?

0
Number of pen color to be used ?
1
Is above information correct ?

YES
```

Plot shown below overlayed curve.

```
Beep will sound when plot done, then press CONTINUE
File name where subroutines are stored ?

LANDER: INTERNAL
Is data medium placed in device INTERNAL
?

YES
Is program medium placed in device ?
YES
Enter convergence coefficient (e.g. 0.005,.001)

.0001
Initial estimate for parameter # 1
?

1000
Initial estimate for parameter # 2
?

-1
Initial estimate for parameter # 3
?

-.2
Initial estimate for parameter # 4
?

1200
Initial estimate for parameter # 5
?

-5.8
Initial estimate for parameter # 6
?

1826
Initial estimate for parameter # 7
?

-16
Initial estimate for parameter # 8
?

.046
Is the above information correct?

YES
*********************************************************************************

Delta(Convergence criteria)=  .0001

THE INITIAL VALUES OF PARAMETERS ARE :
           PARAMETER 1 = 1000
           PARAMETER 2 =-1
           PARAMETER 3 =-.2
           PARAMETER 4 = 1200
           PARAMETER 5 =-5.8
           PARAMETER 6 = 1826
           PARAMETER 7 =-16
           PARAMETER 8 = .046


Would you like to see every iteration ?
YES
```

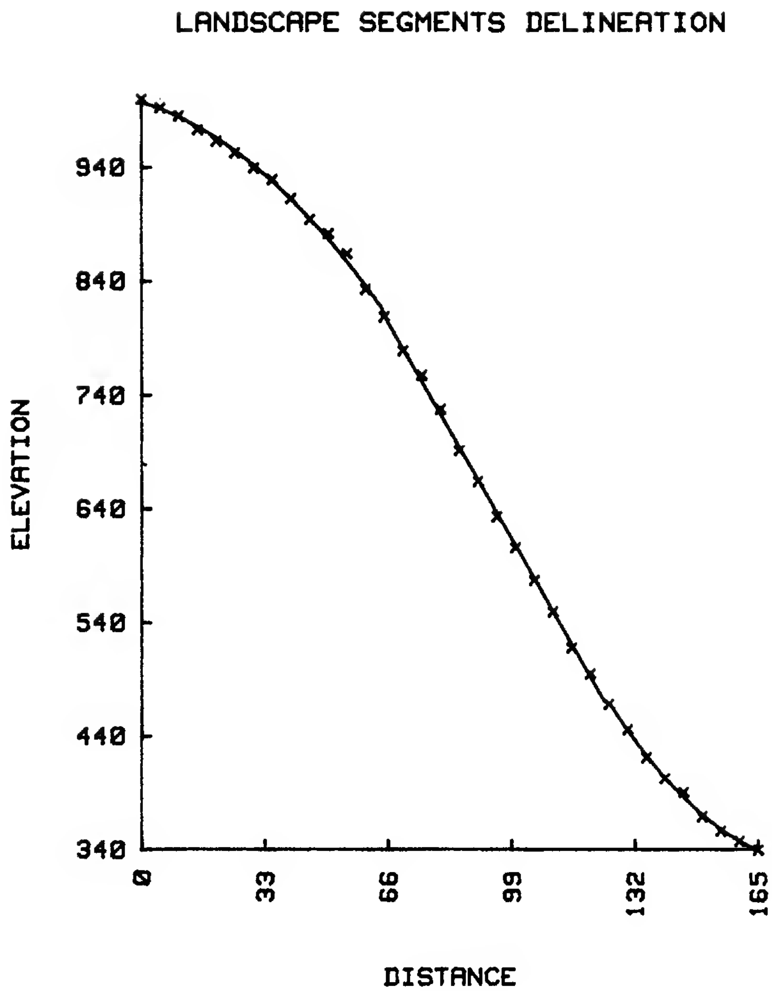
Supply initial estimates.

```
Calcs. may be lengthy. A beep will sound when done. Press 'S' key to ABORT!
ITERATION                 ESTIMATED PARAMETER VALUES                 S.S.RESIDUALS


Calculations may be quite time consuming. A beep will sound when completed
     0      1000.00000        -1.00000         -.20000       1200.00000
          -5.80000     1826.00000      -16.00000          .04600   1693553.53
     1       994.39986        -.69611          -.03291       1184.69313
          -5.76883     1796.78585      -16.20243          .04466       339.21
     2       997.48082       -1.00858          -.02807       1184.09329
          -5.76284     1798.11334      -16.22046          .04472       295.84
     3       997.51105       -1.01126          -.02804       1184.08940
          -5.76280     1807.15342      -16.34364          .04514       295.74
     4       997.51107       -1.01126          -.02804       1184.08939
          -5.76280     1823.35608      -16.56441          .04588       295.65
     5       997.51107       -1.01126          -.02804       1184.08939
          -5.76280     1826.81849      -16.61159          .04604       295.65
 DONE!!!!
```

First eight values per line are the estimated coefficients. Last is sum of squared residuals.

```
******************************************************************************************
THE ESTIMATED PARAMETER VALUES AFTER 5 ITERATIONS ARE :
          PARAMETER  1=      997.5110714 ( 9.9751107143E+02)
          PARAMETER  2=       -1.0112610 (-1.0112609889E+00)
          PARAMETER  3=        -.0280357 (-2.8035714287E-02)
          PARAMETER  4=     1184.0893939 ( 1.1840893939E+03)
          PARAMETER  5=       -5.7627972 (-5.7627972028E+00)
          PARAMETER  6=     1826.8938829 ( 1.8268938829E+03)
          PARAMETER  7=      -16.6126168 (-1.6612616824E+01)
          PARAMETER  8=         .0460476 ( 4.6047611520E-02)
******************************************************************************************
           THE INITIAL VALUE OF SUM OF SQUARED RESIDUALS = 1693553.53
           AFTER 5 ITERATIONS THE SUM OF SQUARED RESIDUALS= 295.649036151
           APPROXIMATE STANDARD ERROR FROM SQUARED RESIDUALS= 3.37210865409
Plot regression curve on present graph ?
YES
Same pen color ?
YES
```

Plot curve or graph.



Like to change initial estimates and/or function ?

NO
Are confidence intervals on parameters desired ?

NO

Residual analysis and/or prediction?

YES

Print out residuals?

YES

TABLE OF RESIDUALS

OBS#	OBSERVED Y	PREDICTED Y	RESIDUAL	STANDARDIZED RESIDUAL	SIGNIF.
1	1000.00000	997.51107	2.48893	.73809	
2	992.40000	991.75387	.64613	.19161	
3	985.40000	984.59489	.80511	.23876	
4	973.30000	976.03412	-2.73412	-.81080	
5	963.10000	966.07157	-2.97157	-.88122	
6	952.90000	954.70723	-1.80723	-.53593	
7	939.60000	941.94110	-2.34110	-.69425	
8	929.40000	927.77319	1.62681	.48243	
9	912.90000	912.20349	.69651	.20655	
10	894.50000	895.23201	-.73201	-.21708	
11	881.80000	876.85874	4.94126	1.46533	
12	864.00000	857.08368	6.91632	2.05104	**
13	832.90000	835.90684	-3.00684	-.89168	
14	808.80000	813.32821	-4.52821	-1.34284	
15	779.00000	780.69359	-1.69359	-.50223	
16	757.40000	751.87960	5.52040	1.63708	
17	727.60000	723.06562	4.53438	1.34467	
18	691.40000	694.25163	-2.85163	-.84565	
19	664.10000	665.43765	-1.33765	-.39668	
20	633.00000	636.62366	-3.62366	-1.07460	
21	605.70000	607.80967	-2.10967	-.62562	
22	577.10000	578.99569	-1.89569	-.56217	
23	549.80000	550.18170	-.38170	-.11319	
24	518.00000	521.36772	-3.36772	-.99870	
25	495.10000	492.55373	2.54627	.75510	
26	468.40000	463.73974	4.66026	1.38200	
27	446.20000	445.45833	.74167	.21994	
28	421.40000	423.40833	-2.00833	-.59557	
29	403.00000	403.66071	-.66071	-.19594	
30	390.90000	386.21548	4.68452	1.38920	
31	369.30000	371.07262	-1.77262	-.52567	
32	356.60000	358.23214	-1.63214	-.48401	
33	347.70000	347.69405	.00595	.00177	
34	340.10000	339.45833	.64167	.19029	

Durbin-Watson Statistic: 1.51322482175

Test statistic for autocorrelation of residuals
Special tables are necessary.

Residual plots?

YES

Plot on CRT?

Residual plots

NO

Plot on external plotter

Plotter identifier string (CONT if 'HPGL') ?

Plotter select code, Bus # = (defaults are 7,5)

Residual plot option no. =

1

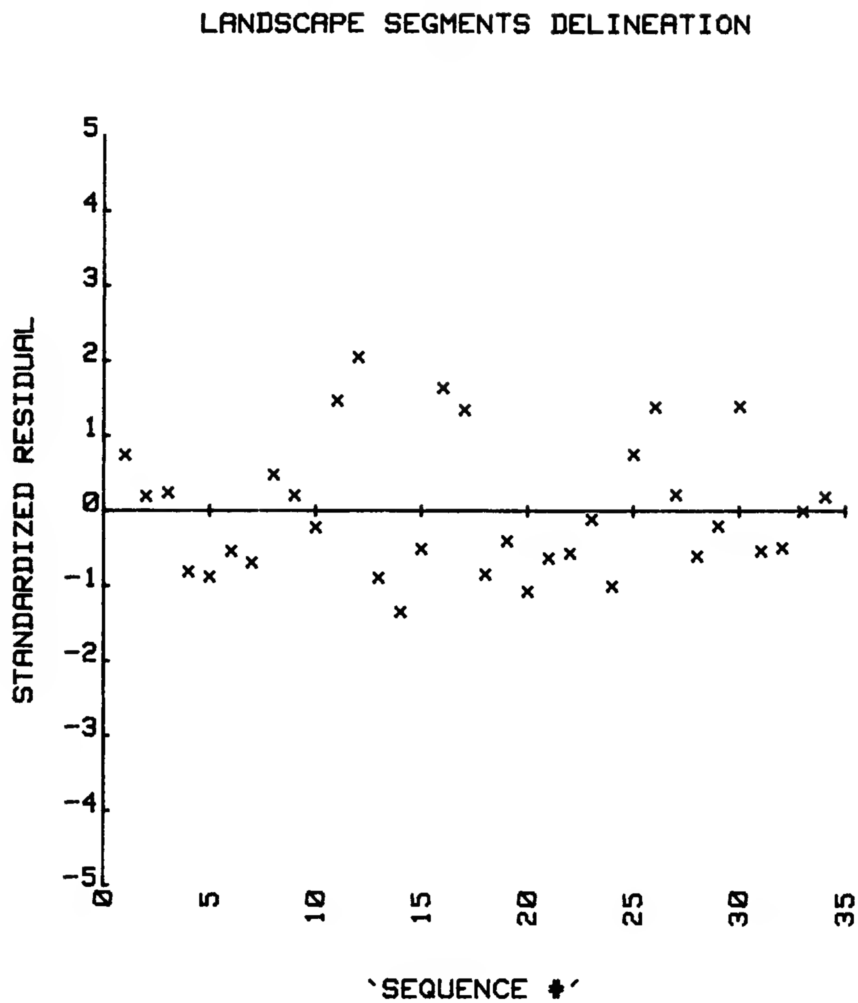
Plot residuals vs time square

For plotting, X-min = ?

0

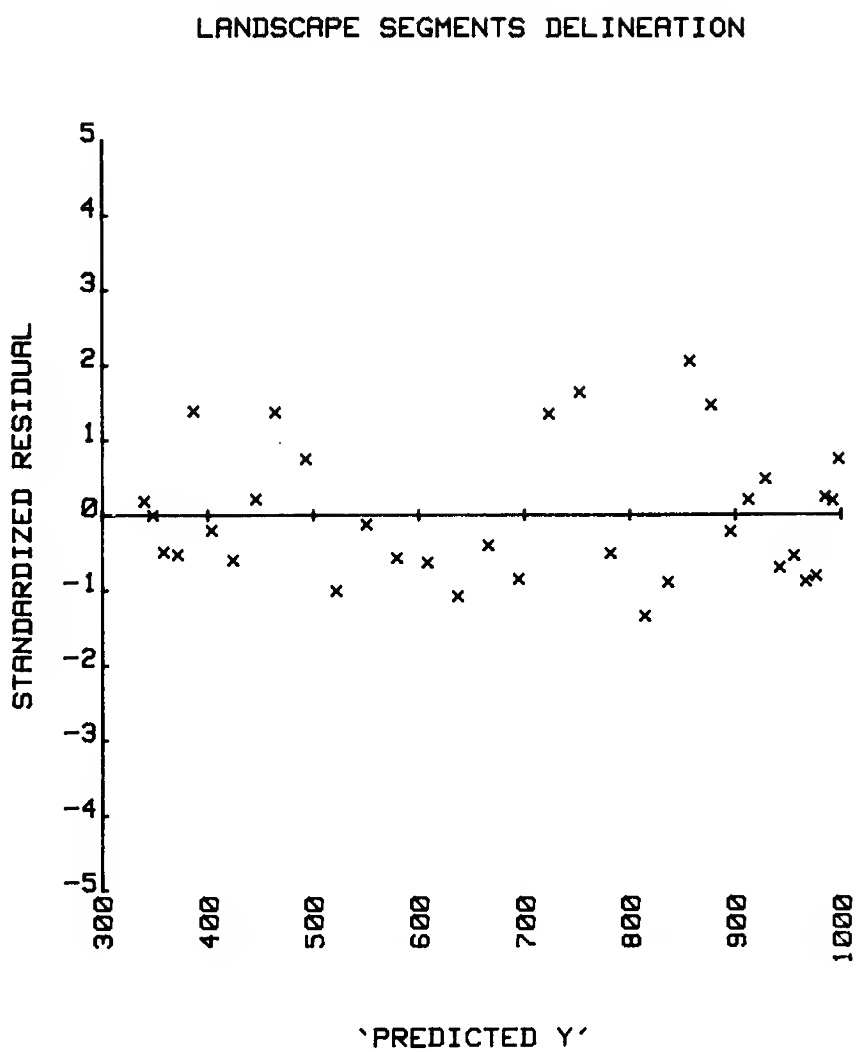
For plotting, X-max = ?

```
35
Distance between X-ticks = ?

5
# of decimals for labelling X-axis (<=7) = ?

0
Number of pen color to be used ?
1
Is above information correct?

YES
```




```
Residual plots ?
YES
Would you like to plot on CRT ?
NO
Plotter identifier string (CONT if 'HPGL') ?
Plotter select code, bus # (defaults are 7,5) ?
```

```
Residual plot option no. = ?

2                                                    Plot residuals vs predicted Y
For plotting, X-min = ?

300
For plotting, X-max = ?

1000
Distance between X-ticks = ?

100
# of decimals for labelling X-axis (<=7) = ?

0
Number of pen color to be used ?
1
Is above information correct?

YES
```




```
Residual plots ?

NO                                                   Exit residual routine.
Option number = ?

7                                                    Return to BSDM
```

Example 6: Standard Nonlinear Regression

In this example, standard nonlinear models are fit to the data from Example 4.

```
Are you going to use user defined transformation
or do Non-linear regression ? (Y/N)

NO
Are you using an HPIB Printer?

YES
Enter select code, bus address (if 7,1 press CONT) ?

********************************************************************************
*                              DATA MANIPULATION                               *
********************************************************************************

Enter DATA TYPE:

1                                                    Raw data (data type required)
Mode number = ?                                      From mass storage

2
Is data stored on the program's scratch file (DATA)?

YES                                                  Previously stored on program's scratch
                                                     file called DATA.

                     URINE/BLOOD CONCENTRATION

Data file name: DATA

Data type is:   Raw data

Number of observations:    25
Number of variables:        2

Variable names:
   1. TIME(HR)                                       Same data set which we used for nonlinear
   2. BLD.CONT                                       regression.

Subfiles:  NONE

SELECT ANY KEY
                                                     Select special function key labeled-LIST
Option number = ?

1                                                    List all the data
Enter method for listing data:

3                                                    In tabular form
```

URINE/BLOOD CONCENTRATION

Data type is: Raw data

OBS#	Variable # 1 (TIME(HR))	Variable # 2 (BLD.CONT)
1	4.25000	1165.70000
2	7.50000	851.00000
3	10.80000	523.00000
4	12.00000	365.00000
5	16.00000	294.00000
6	23.80000	170.00000
7	27.80000	60.00000
8	35.30000	81.00000
9	38.30000	20.00000
10	45.30000	45.00000
11	51.30000	27.00000
12	54.20000	37.00000
13	59.80000	31.00000
14	64.25000	26.00000
15	69.50000	36.00000
16	78.20000	18.00000
17	90.20000	10.00000
18	100.00000	8.20000
19	105.00000	13.40000
20	108.00000	17.40000
21	114.00000	8.00000
22	120.00000	4.00000
23	130.00000	6.70000
24	142.00000	6.70000
25	154.00000	5.80000

Option number = ?

0 — Exit List routine.

SELECT ANY KEY

Select special function key labeled-ADV STAT
Remove BSDM medium.
Insert regression medium.

Option number = ?

5 — Select standard non-linear regression modes.

Number of the regression model = ?

3 — Mixed exponential of form:
Y = A*Exp(B*X) + C*Exp(D*X)

Should fitted model include intercept term ?

NO — Note: In the non-linear regression example we specified 3 exponential terms.

Number of the dependent variable = ?

2 — Y = blood count

Number of the independent variable = ?

1 — X = time in hours

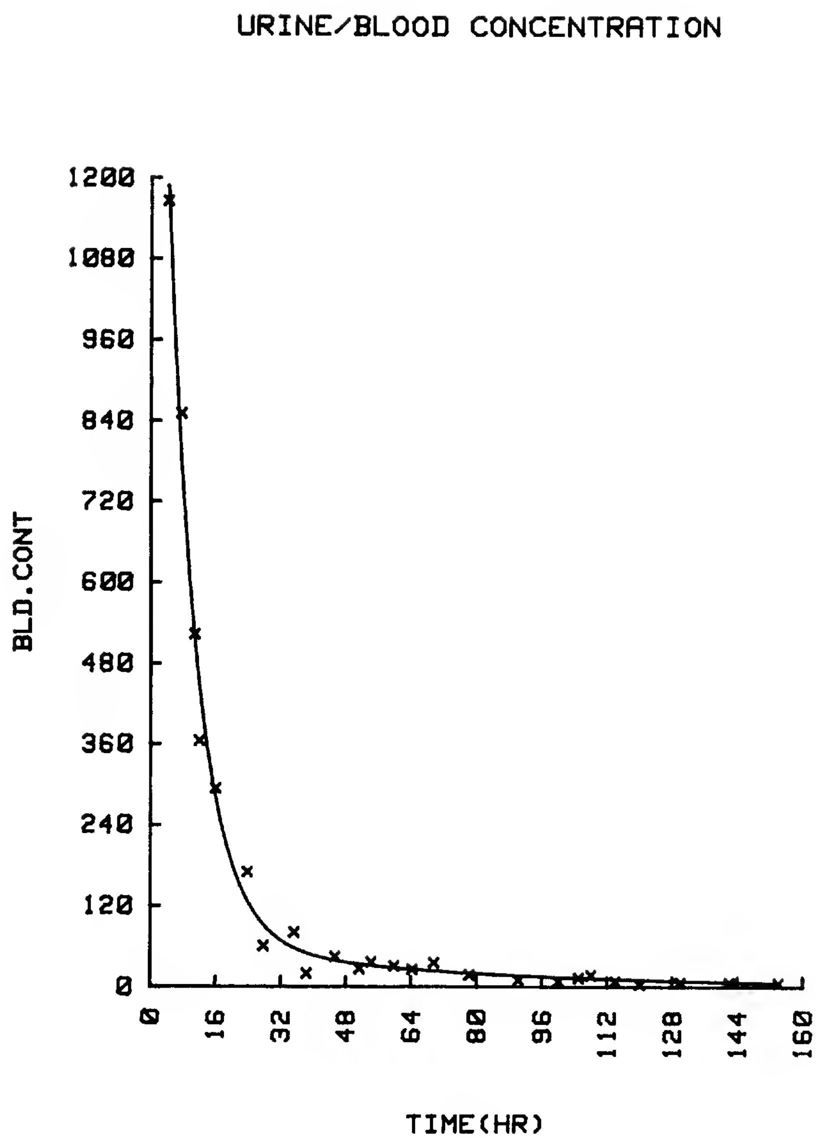
Is above information correct?

YES — Displayed on CRT. It is correct.

```
*****************************************************************************
REGRESSION MODELING ON DATA SET:
                           URINE/BLOOD CONCENTRATION
*****************************************************************************

--where:  Dependent variable = (2)BLD.CONT
          Independent variable = (1)TIME(HR)

THE STANDARD NON-LINEAR REGRESSION MODEL SELECTED = Y=A*EXP(B*X)+C*EXP(D*X)


Is a plot of the regression desired?

YES                                                   Like to see a plot
Plot on CRT

NO                                                    But not on CRT.
Plotter identifier string (CONT if 'HPGL') ?

Plotter select code,Bus# =(defaults are 7,5) ?        On an external plotter at 7,5
X-min = ?

0
X-max =

160                                                   Specify plotting limits.
Y-min = ?

0
Y-max = ?

1200
Y-axis crosses X-axis at X = ?

0
X-axis crosses Y-axis at Y = ?

0
Distance between X-ticks = ?

16
Distance between Y-tic =

120
# of decimals for labelling X-axis((=7) = ?

0
# of decimals for labelling Y-axis = ?

0
Number of pen color to be used ?
1
Is above information correct ?

YES
Is plotter ready ?

YES
Are the values of the initial estimates proper?       As shown on CRT and printed out below.

YES
```

```
********************************************************************************

Delta(Convergence criteria)=  .05

THE INITIAL VALUES OF PARAMETERS ARE :
          PARAMETER 1 = 334.489319026
          PARAMETER 2 =-3.26684362156E-02
          PARAMETER 3 = 33.4489319026
          PARAMETER 4 =-1.63342181078E-02


Calcs. may be lengthy. A beep will sound when done. Press 'NO' key to INTERRUPT
!
CALCULATIONS STARTED ON  0 / 0  AT  0  0 : 0


ITERATION              ESTIMATED PARAMETER VALUES               S.S.RESIDUALS
                 A                B              C              D
    0      334.48932          -.03267       33.44893       -.01633 1137486.3294
    1      767.19521          -.09251      211.31532       -.03542  330889.6618
    2     1593.74645          -.18542      335.76599       -.03753   82374.8400
    3     1854.77884          -.13293      214.04524       -.03737   39473.4982
    4     1974.36951          -.14275      120.85302       -.02902   17809.6312
    5     2008.32686          -.13849       78.90472       -.01868   17060.8039
    6     2003.21510          -.13699       73.85445       -.01677   16989.6350
 DONE!!!!
********************************************************************************
THE ESTIMATED PARAMETER VALUES AFTER 6 ITERATIONS ARE :
           PARAMETER  1=     2002.9416350 ( 2.0029416350E+03)
           PARAMETER  2=        -.1371748 (-1.3717475809E-01)
           PARAMETER  3=       75.2098521 ( 7.5209852103E+01)
           PARAMETER  4=        -.0170887 (-1.7088737873E-02)
********************************************************************************
           THE INITIAL VALUE OF SUM OF SQUARED RESIDUALS = 1137486.32942
           AFTER 6 ITERATIONS THE SUM OF SQUARED RESIDUALS= 16989.1765347
           APPROXIMATE STANDARD ERROR FROM SQUARED RESIDUALS= 28.4430730832
********************************************************************************
Should regression line be plotted on same graph ?
YES
Same pen color ?
YES
```

Note: These results in terms of the sum of squared residuals are very close to the non-linear regression example with two more parameters.



New initial estimates and/or convergence criteria ?

NO Satisfied with results.
Are confidence intervals on parameters desired ?

YES Why not get confidence intervals.
Confidence coefficient for confidence interval on parameters(e.g. 90,95,99)=
95

95 % CONFIDENCE INTERVALS ON PARAMETERS

PARAMETER	ONE-AT-A TIME C.I.		SIMULTANEOUS C.I.	
	LOWER LIMIT	UPPER LIMIT	LOWER LIMIT	UPPER LIMIT
1	1818.4543	2187.4289	1703.9352	2301.9481
2	-.1593	-.1150	-.1731	-.1013
3	-42.6712	193.0909	-115.8451	266.2648
4	-.0433	.0092	-.0596	.0255

****

```
Residual analysis and/or prediction ?

YES
Print out residuals?

YES
```

Residual analysis

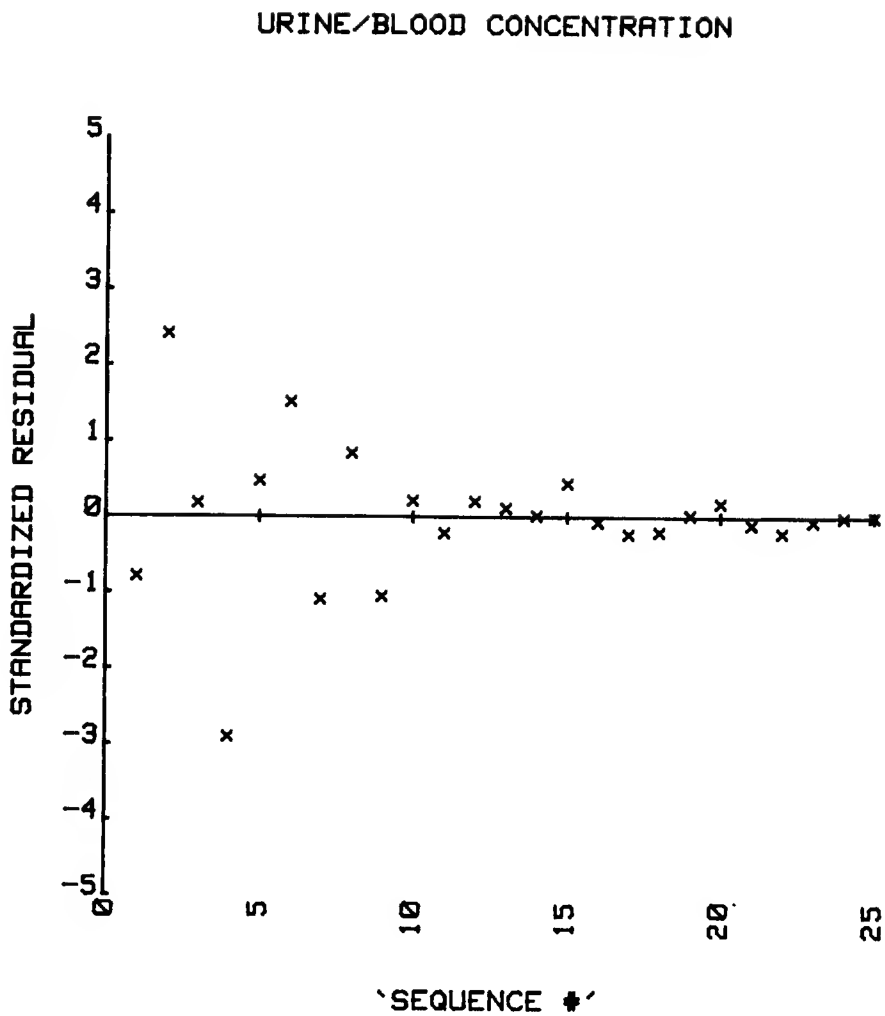
TABLE OF RESIDUALS

OBS#	OBSERVED Y	PREDICTED Y	RESIDUAL	Residual Sy.X STANDARDIZED RESIDUAL	SIGNIF.
1	1165.70000	1188.03381	-22.33381	-.78521	
2	851.00000	782.07774	68.92226	2.42317	**
3	523.00000	517.80218	5.19782	.18274	
4	365.00000	447.43453	-82.43453	-2.89823	**
5	294.00000	280.30783	13.69217	.48139	Note:
6	170.00000	126.60208	43.39792	1.52578	Two large
7	60.00000	90.97711	-30.97711	-1.08909	residuals.
8	81.00000	56.94431	24.05569	.84575	
9	20.00000	49.55743	-29.55743	-1.03918	
10	45.00000	38.68823	6.31177	.22191	
11	27.00000	33.06036	-6.06036	-.21307	
12	37.00000	30.96936	6.03064	.21202	
13	31.00000	27.61708	3.38292	.11894	
14	26.00000	25.38439	.61561	.02164	
15	36.00000	23.07884	12.92116	.45428	
16	18.00000	19.80955	-1.80955	-.06362	
17	10.00000	16.10942	-6.10942	-.21479	
18	8.20000	13.62043	-5.42043	-.19057	
19	13.40000	12.50406	.89594	.03150	
20	17.40000	11.87886	5.52114	.19411	
21	8.00000	10.72091	-2.72091	-.09566	
22	4.00000	9.67600	-5.67600	-.19956	
23	6.70000	8.15597	-1.45597	-.05119	
24	6.70000	6.64379	.05621	.00198	
25	5.80000	5.41199	.38801	.01364	

```
Durbin-Watson Statistic: 2.57642711573


Residual plots?

YES
Plot on CRT?

NO
Plotter identifier string (press CONT if 'HPGL') ?
Plotter select code, Bus # = (defaults are 7,5)

Residual plot option no. = ?
1
For plotting, X-min = ?

0
For plotting, X-max = ?

25
Distance between X-ticks =

5
# of decimals for labelling X-axis ((=7) = ?

0
Number of pen color to be used ?
1
```

Specify limits for residual plot verses sequence #.

```
Is above information correct?

YES
```




```
Residual plots ?
NO                                          Exit residual routine.
Option number = ?

5                                           Standard non-linear regression models.
Number of the regression model = ?          This time with intercept term.

3                                           Mixed exponentials
Should fitted model include intercept term ?

YES                                         This time with an "intercept" term.
Number of the dependent variable = ?        Y=A*EXP(B*X)+C*EXP(D*X)+E

2
Number of the independent variable = ?

1
Is above information correct?

YES
```

```
*********************************************************************************
REGRESSION MODELING ON DATA SET:
                              URINE/BLOOD CONCENTRATION
*********************************************************************************

--where:  Dependent variable = (2)BLD.CONT
          Independent variable = (1)TIME(HR)

THE STANDARD NON-LINEAR REGRESSION MODEL SELECTED = Y=A*EXP(B*X)+C*EXP(D*X)+E

                          IN RADIANS
Is a plot of the regression desired?

NO                                                          No plot this time.
Are the values of the initial estimates proper?

YES
*********************************************************************************

Delta(Convergence criteria)=   .05

THE INITIAL VALUES OF PARAMETERS ARE :
          PARAMETER 1 = 429.352234007
          PARAMETER 2 =-.0428279809939
          PARAMETER 3 = 42.9352234007
          PARAMETER 4 =-.021413990497
          PARAMETER 5 = 3.92



Would you like a hard copy of every iteration ?

YES
Calcs. may be lengthy. A beep will sound when done. Press 'NO' key to INTERRUPT
!
CALCULATIONS STARTED ON  0 / 0  AT  0  0 : 0


ITERATION                ESTIMATED PARAMETER VALUES               S.S.RESIDUALS


     0       429.35223        -.04283        42.93522         -.02141
            3.92000                                                  909776.40446
     1       885.18106        -.09863        211.11731        -.09760
            36.47985                                                 276063.40271
     2      1286.23842        -.12132        604.50758        -.17961
            20.86657                                                  50454.00149
     3      1243.85321        -.10655        824.55159        -.18667
            18.44537                                                  18850.77414
 DONE!!!!
*********************************************************************************
THE ESTIMATED PARAMETER VALUES AFTER 3 ITERATIONS ARE :
           PARAMETER  1=      1218.8442529   (1.2188442529E+03)
           PARAMETER  2=         -.1063836   (-1.0638358605E-01)
           PARAMETER  3=       860.7805920    8.6078059213E+02)
           PARAMETER  4=         -.1848468  (-1.8484679940E-01)
           PARAMETER  5=        17.7594408   ( 1.7759440852E+01)
*********************************************************************************
           THE INITIAL VALUE OF SUM OF SQUARED RESIDUALS = 909776.404501 Not as good
           AFTER 3 ITERATIONS THE SUM OF SQUARED RESIDUALS= 18803.5777771  as before.
           APPROXIMATE STANDARD ERROR FROM SQUARED RESIDUALS= 30.6623366503
*********************************************************************************
New initial estimates and/or convergence criteria ?

NO
Are confidence intervals on parameters desired ?
```

YES
Confidence coefficient for confidence interval on parameters(e.g. 90,95,99)=

95

95 % CONFIDENCE INTERVALS ON PARAMETERS

PARAMETER	ONE-AT-A TIME C.I.		SIMULTANEOUS C.I.	
	LOWER LIMIT	UPPER LIMIT	LOWER LIMIT	UPPER LIMIT
1	631.4045	1806.2840	182.0381	2255.6504
2	-.1322	-.0806	-.1519	-.0609
3	228.0449	1493.5163	-255.9710	1977.5322
4	-.3155	-.0542	-.4154	.0457
5	1.8285	33.6904	-10.3579	45.8768

****
Residual analysis and/or prediction ?

YES
Print out residuals?

YES

TABLE OF RESIDUALS

OBS#	OBSERVED Y	PREDICTED Y	RESIDUAL	STANDARDIZED RESIDUAL	SIGNIF.
1	1165.70000	1185.65897	-19.95897	-.65093	
2	851.00000	781.76841	69.23159	2.25787	**
3	523.00000	521.01910	1.98090	.06460	
4	365.00000	451.45738	-86.45738	-2.81966	**
5	294.00000	284.66099	9.33901	.30458	
6	170.00000	125.23916	44.76084	1.45980	
7	60.00000	86.12756	-26.12756	-.85211	
8	81.00000	47.53330	33.46670	1.09146	
9	20.00000	39.20569	-19.20569	-.62636	
10	45.00000	27.79845	17.20155	.56100	
11	27.00000	23.02251	3.97749	.12972	
12	37.00000	21.61557	15.38443	.50174	
13	31.00000	19.87723	11.12277	.36275	
14	26.00000	19.07606	6.92394	.22581	
15	36.00000	18.51147	17.48853	.57036	
16	18.00000	18.05704	-.05704	-.00186	
17	10.00000	17.84239	-7.84239	-.25577	
18	8.20000	17.78867	-9.58867	-.31272	
19	13.40000	17.77661	-4.37661	-.14274	
20	17.40000	17.77192	-.37192	-.01213	
21	8.00000	17.76603	-9.76603	-.31850	
22	4.00000	17.76292	-13.76292	-.44885	
23	6.70000	17.76064	-11.06064	-.36072	
24	6.70000	17.75978	-11.05978	-.36070	
25	5.80000	17.75953	-11.95953	-.39004	

Durbin Watson Statistic: 2.36053763341

Residual plots?

NO
Option number = ?

7 Return to BSDM

Statistical Graphics

General Information

Object of Program

This group of nine programs has been developed to allow you to quickly get a graphical representation of your data with a minimum number of questions on the CRT screen or an HP-IB Peripheral Plotter.

Because of the length of the programs, two discs are used to hold the Statistical Graphics Routines.

The entry to every program requires that you specify only the variables to be used and how subfiles are to be treated if they exist. From here on, all plotting parameters are determined by the program, and a plot may be constructed immediately by selecting the plot option from the plotting characteristics menu.

Once the data has been specified, you have the option of changing nearly all the plotting parameters in order to construct a more personalized plot. This is done by selecting the option from the plotting characteristics menu.

Any time new variables are defined by selecting the "RESTART" option from the menu, all previous parameters that had been defined are reset to a default value for this particular data set. In order to save the plotting characteristics you have specified, select the store option available in the menu. This stores the plotting characteristics out on another file of you choice. Then, after you select the restart option, you can retrieve these characteristics by selecting the load option.

Special Considerations

1. Every time you select a graph type, the CRT is declared as the standard plotting device. This unit may be changed by selecting the "Select Plotter" option from the plotting characteristics menu.
2. Every program begins its execution by reloading the data contained in the "DATA" file. In the case of the NORMAL PROBABILITY PLOT and WEIBULL PROBABILITY PLOT, the file "DATA" is reloaded every time the "RESTART" option is selected.

3. The "RESTART" option always initializes the plotting parameters to default values as follows:
   - The axes labels default to the name of the variable being plotted.
   - The graph title contains the first 33 characters of the data set title. If a subfile is declared, the graph title preceded by the 10-character name given to the subfile.
   - The plotting symbol is a plus sign.
   - Pen numbers are set to 1.
   - The axis parameter is wide enough to contain the data set and has 10 equally spaced tic marks with every second tic mark being labeled.
   - In the special case of the log axis, only complete cycles are plotted on a log scale might be scaled so that it fits in an entire cycle.
   - The graphics device used for plotting is reset to CRT.

Note

After selecting the "OVERLAY" option, new data may be plotted on the previously constructed graph. But, the default values will be in effect for pen number and symbol. These may be changed by selecting the "SELECT PLOTTER" and "SELECT PEN NUMBER" options from the Plotting Characteristics menu.

4. Whenever the program identifies an incorrect response, the question is asked again, until the correct response is given.
5. Most plotting symbols are centered on top of the point they are designating. For some special characters, like the period and comma, the symbols are plotted in a lower position.
6. The graphics programs only allow up to six decimal places for labeling the axis tic marks. For data that would need more, it is suggested that it be transformed.
7. Each program handles missing values in a different way. See the individual programs for details.
8. When asking for labeling information, an error 18 will occur if the label is too long. To recover, shorten the label and re-enter it.
9. Do no press the "RUN" or "SHIFT-PAUSE" (RESET) keys unless it is necessary. The "RUN" key erases all variables, and RESET may erase memory.
10. To prevent a graph segment from being plotted, assign a pen number of -1 for the CRT or 0 for an external plotter to that segment using the select pen numbers option.

Note

Statistical Graphics may be entered from any of the other Statistics packages by selecting the Advanced Statistics option. Once in the Statistical Graphics package, select the type of plot you wish to do from the menu provided.

Common Plotting Characteristics

The following options are available for all nine of the plots, so their description and operation are explained in this section. There are slight deviations in the way some of these options work for the different plots. These differences are explained in the sections that describe each plot. It is recommended that you read through the section for the particular plot you wish to do before using the program. Not all plotting characteristics can be changed in each program.

RESTART

When this option is selected, all plotting parameters for the data set are reset to the default values which the program has determined for the data set. At this time a new variable to be plotted may be selected.

PLOT

This option plots the variable(s) being considered. The plot will be done on the CRT if no other device has been specified. If you have not specified any plotting parameters, the ones determined by the program are used, otherwise, the plotting characteristics you specified are used. You may choose whether or not to connect the points on most of the graphs, and whether or not to put grid lines on the graph.

X-AXIS

This option allows you to designate the scale for the x-axis. You determine the minimum x value, the maximum y value, the distance between the tic marks on the axis, and how many places after the decimal point you want printed. Since complete cycles on the x-axis are required by the semi-log, log-log, normal and Weibull plots, this option may not be used in those routines.

Y-AXIS

This option allows you to designate the scale for the y-axis. You determine the minimum y value, the maximum y value, the distance between the tic marks on the axis, and how many places after the decimal point you want printed. Since full cycles are required on the y-axis by the log-log and Weibull plots, this option may not be used in those routines.

LABELS

This option allows you to change the labels of your graph. You have an opportunity to change the x-axis label, the y-axis label, and the title of the graph.

SYMBOLS

This option allows you to change the symbol used to designate the points on the graph. If you do not want any symbol use a blank which is designated by " ".

Dump Graphics On CRT

This option prints the most recent CRT graph on the printer. This option may be used **only** if your printer has graphics capabilities (e.g. 2671G, 2631G).

SELECT PLOTTER

This option allows you to select the plotting device on which you wish to have the plot drawn. You may have the plot done on the CRT or an external plotter. You will need to input the select and bus codes. You will also need to input a plotter identification string.

SELECT PEN COLOR

This option allows you to select the pen number you wish to use for plotting your graph. The pen number used may be changed for axes and numeric labels, grid lines, labels and points.

OVERLAY

This option, when available, allows you to add another plot of the same type with new variables on the previously constructed plot. The plotting limits will remain as you have specified.

STORE

This option allows you to store the plotting characteristics that you have specified so that they may be retrieved at a later time. To do this you need to specify a file name and where you wish to store the information.

LOAD

This option allows you to retrieve the plotting characteristics that were stored previously for this type of plot. You need to specify the name of the file and where it was stored. The program will then list the stored plotting characteristics.

RETURN

This option returns the program to the main STATISTICAL GRAPHICS MENU.

Time Plot

Object of Program

This program plots any variable in increasing units of time or sequence number. This plot is useful in determining the effect that time/sequence may have on a variable. The program allows the initial time to begin the plotting and the time period between points to be set by selecting the "X AXIS" option. If the plot option is selected first, the program defaults to a starting time of 1 and time increment period of 1.

Special Considerations

1. Missing values are not plotted. The value at this time period is left blank.
2. When doing an overlay of the data, the initial time and time increments are 1 unless changed by selecting the x-axis option. Once the values have been changed, they retain the new values until they are changed again.

Special Plotting Characteristics

X-AXIS

This option allows you to determine the scale for the time axis. You need to specify the minimum and maximum time values, and the distance between tic marks. In addition, you need to specify the initial time for beginning series, the point in time that the plotting begins, and time increments between points, how much time passes between each plotted point.

OVERLAY

This option allows you to plot another variable over an already contructed graph.

References

1. EXPLORATORY DATA ANALYSIS, John W. Tukey; 1977; Addison Wesley.
2. "A Review of Some Smoothing and Forecasting Techniques", T. J. Boardman and M.C. Bryson, Journal of Quality Technology, Volume 10, Number 1, January, 1978.

Histogram

Object of Program

This program creates a histogram with up to forty cells. For every data set, the sample mean, the sample variance, the number of cases used to calculate them, and the cell statistics will be printed.

Different histograms may be created by specifying the number of cells to be used, and the cell locations, or by specifying the number of cells, the location of the first cell, and the cell width. These specifications may be given by selecting the "CELL LIMIT" option from the Plotting Characteristics menu.

A normal curve overlay and the corresponding Chi-squared goodness-of-fit statistic may be obtained by selecting the "NORMAL CURVE OVERLAY" option from the Plotting Characteristics menu.

Special Considerations

1. Missing values are not considered in any calculation, and are not considered in constructing any cell.
2. A maximum number of forty cells may be obtained.
3. At least four cells are needed to perform a chi-squared goodness-of-fit test.

Special Plotting Characteristics

CELL LIMITS

This option allows you to specify the cell size for the histogram. There are two ways of doing this:

1. Enter the number of cells (greater than 1 but not more than 40) and Enter the minimum cell value and the maximum cell value that should be used.
2. Enter the number of cells (greater than 1 but less than 40), and Enter the mimumum cell value and the width of the cell.

The program will then give you a list of the number of cells, their minimum and maximum bounds, and the number of observations in each cell.

NORMAL CURVE OVERLAY

This option does a chi-square goodness-of-fit test of the data. In order to do this at least four cells must be specified; if four cells have not been specified, an error will be printed. The descriptive statistics for each cell will be printed. The contributions to the chi-squared statistics are added together to get the final value. The cells on the tails are collapsed together until an expected frequency of at least three and less than seven is found, and then the contribution is calculated. If, after collapsing the end cells to get high enough frequencies, the number of terms in the contribution of the chi-squared value go below four then another error will be printed.

Once this is done the normal curve for the desired plot is plotted over the histogram.

Methods and Formulae

X_i = ith observation of the selected variable that is not a missing value

N = number of valid observations

$$\bar{X} = \sum_{i=1}^{N} X_i/N$$

$$\text{Variance} = \frac{\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2 /N}{N-1}$$

Normal Curve overlay =

$$\frac{100*(\text{Cell width})*(\text{EXP}((X-\bar{X})^2/(2*\text{Variance}))}{\sqrt{2\pi*\text{Variance}}}$$

$$\chi^2_{df} = \sum_{i=1}^{\#\text{ cells}} \frac{(\text{Observed frequency in cell i} - \text{Expected frequency in cell i})^2}{(\text{Expected frequency in cell i})}$$

df = (# of cells) − 3; because 1 degree of freedom is lost for number of cells, 1 for the estimated mean, and one for the estimated variance.

The expected frequency of cell i = area under the normal curve overlay which would fall in cell i is calculated by determining the left side of the cell i(A), and the right side of the cell i(B) and finding

$$\Phi\left(\frac{B - \text{Xbar}}{\text{standard deviation}}\right) - \Phi\left(\frac{A - \text{Xbar}}{\text{standard deviation}}\right)$$

Then use the following approximation for the area between A and B in a standard normal.

$\Phi(X) = 1 - Z(X)\ (b_1t + b_2t^2 + b_3t^3 + b_4t^4 + b_5t^5) + E(X)$ where $|E(X)| < 7.5*10^{-8}$
$t = (1 + .231649X)^{-1}$
$b_1 = .31938153$
$b_2 = -.35656378$
$b_3 = 1.781477939$
$b_4 = 1.821255978$
$b_5 = 1.330274429$
for $X>0$
and $1 - \Phi(|X|)$ for $X \leq 0$
$Z(X) = \exp(-x^2/2)/\sqrt{2\pi}$

To calculate the right tailed probability value associated with the Chi Square value we use

$P(X^2\nu > \text{calculated value}) =$

$$1 - \left\{\left[\frac{\chi^2}{2}\exp\left(-\chi^2/2\right)\right] / \Gamma\left(\left(\nu+2\right)/2\right)\right\} *C$$

$$C = 1 + \sum_{R=1}^{\infty} \frac{\chi^{2R}}{(\nu+2)\ (\nu+4)\ldots(\nu+2R)}$$

where X^2 is the calculated value
ν is the degree(s) of freedom
γ (.) is the standard gamma function $\gamma(.5) = .88626925$
The sum is calculated until the percentage of change between two consecutive sums is less than .000001 or $R = 40$.

The number of cells being used defaults to the value given by the closest integer of the function:

$[1 + (3.3\log_{10}$ (Number of valid observations))]

References

1. An Introduction to Statistical Methods and Data Analysis, Lyman Ott; 1977; Wadsworth.
2. Statistics for Modern Business Decisions, Second Edition, Lawrence Lapen; 1978; Harcourt, Brace, Jovanovich.
3. Statistical Analysis for Decision Making, Second Edition, Morris Hamburg; 1977; Harcourt, Brace, Jovanovich.
4. Fundamental Statistics for Business and Economics, Fourth Edition; Neter, Wasserman, and Whitmore; 1973; Allyn and Bacon.
5. Handbook of Mathematical Functions, Abramowitz, Stegun; Fifth Printing; 1965 Dover Publications.

Normal Probability Plot

Object of Program

This program creates normal probability paper, orders the data, and then plots the data on the paper. This plot may be used to indicate if the data set may have come from a normal distribution. If a straight line can be made to fit the plotted points, then the data may come from a normal distribution.

Special Considerations

1. Missing values are eliminated from the data, which effectively makes the data set one smaller for each missing value.
2. When plotting more than a hundred points, it is suggested that the period be used as the plotting symbol. This allows for a more even line. Note that the period is plotted lower than the actual value of the point.
3. A maximum of 999 points may be plotted on the graph with the empirical distribution used by the program.

Special Plotting Characteristics

LABELS

This option allows you to change the labels for the y-axis and the title, but not the x-axis.

OVERLAY

This option allows you to plot the normal probability of another variable over the already existing graph.

Methods and Formulae

Empirical Distribution Function (EDF)

X_i is the i sorted value in the data set. i can go from 1 to N. N is the number of non-missing values in the data set. EDF(Xi) = i/(N + 1)

Cumulative distribution function (CDF) for plotting and scaling the X axis is done by determining the EDF(Xi) and then determining X_p.

$$X_p = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}$$

where $t = \sqrt{\log_e(1/(EDF(X_i))^2}$

$c_0 = 2.515517$
$c_1 = .802853$
$c_2 = .010328$
$d_1 = 1.432788$
$d_2 = .189269$
$d_3 = .001308$

References

1. Probability Plots for Decision Making, James R. King; 1971; Industrial Press.
2. "Weibull Probability Papers", Wayne Nelson and Vernon C. Thompson, Journal of Quality Technology; Volumn 3, Number 2, April 1971.

Weibull Probability Plot

Object of Program

This program creates Weibull probability paper, orders and then plots the data. The number of cycles used to plot the data is determined by the data.

If the plotted data appears to lie on a straight line, the data may come from a Weibull distribution. No attempt is made in the program or on the paper to estimate the parameters of the Weibull distribution.

Special Considerations

1. Missing values are eliminated from the data, which effectively makes the data set 1 smaller for each missing value.
2. When more than a hundred points are plotted, it is suggested that the period be used as the plotting symbol. This allows for a more even, narrower line. Note that the period is plotted lower than the actual value of the point.
3. A maximum of 999 points may be plotted on the graph with the empirical distribution used by the program.
4. All data used by this program must be positive. The data is checked and a message is printed if any zero or negative data is found.

Methods and Formulae

Empirical Distribution Function (EDF)

X_i is the ith sorted value in the data set. i can go from 1 to N where N is the number of non-missing values in the data set.

$$\mathrm{EDF}(X_i) = i/(N+1)$$

$$\text{Percent Failure} = \log_e\left(\log_e\left(\frac{1}{1-\mathrm{EDF}(X_i)}\right)\right)$$

Scattergram

Object of Program

This program plots points on a graph according to the two variables you specify. The plot is useful in determining if there is any relationship between two variables.

Special Considerations

For any point where either the X or Y coordinate is missing, the point is not plotted.

Semi-Log Plot

Object of Program

This program plots points on a graph where each X value is plotted on a log scale, and each Y value is plotted on a normal scale. The number of cycles used on the X axis is determined by the program.

This plot is useful in determining if any relationship between an untransformed Y variable and a log-transformed X variable exists.

Special Considerations

1. For any point where either the X or Y coordinate is missing, the point is not plotted.
2. All data used for the X variable must be greater than zero.

Log-Log Plot

Object of Program

This program plots points on a graph where both the X and Y axes take on log values. The number of cycles used by both axes are determined by the program.

The plot of the points is useful in determining if any relationship exists between log-transformed X and Y variables.

Special Considerations

1. For any point where either the X or Y coordinate is missing, the point is not plotted.
2. All data specified for this program must be positive.

References

1. Exploratory Data Analysis, John W. Tukey; 1977; Addison Wesley.
2. The Statistical Analysis of Experimental Data, John Mandel; Interscience.

3D Plot

Object of Program

This program constructs and draws points in a simulated three-dimensional graph. The axes may be rotated and tilted to see relationships between the data better. An effective XY scatter-plot may be obtained by tilting the axes 90 degrees. The program looks best when rotation and tilt are between 20 and 70 degrees. At more extreme angles, labeling problems may occur. You may correct some of these problems by adjusting the axis so that the number of tic marks labeled are fewer, and so that axes labels are shorter.

Special Considerations

1. For any point where either the X, Y or Z value is missing, the point is not plotted.
2. For long axes titles and various rotation and tilt combinations, the axes titles may overlap, or not be entirely plotted.

Special Plotting Characteristics

PLOT

This option plots the three variables that were specified. You need to input the angle (in degrees) of rotation about the z-axis between zero and ninety degrees, and the angle, between zero and ninety of elevation, which is the angle between the line drawn from the origin of the axes and the XY plane.

Z-AXIS

This option allows you to designate the scale for the z-axis. It works the same as the options for the X and Y-axis.

Methods and Formulae

Mapping from the third dimension to the two dimensions of the plotting device uses the following method.

Given any point (X,Y,Z) we map to the point (A,B) by letting

$$A = \frac{(X - Xmin)}{(Xmax - Xmin)}(COS(Rotation)) + \frac{(Y - Ymin)}{(Ymax - Ymin)}(-SIN(Rotation))$$

and

$$B = \frac{(X - Xmin)[(COS^2 (Rotation) - 1)(TAN(Tilt/2))]}{(Xmax - Xmin)}$$

$$+ \frac{(Y - Ymin)[(SIN^2 (Rotation) - 1) (TAN(Tilt/2))]}{(Ymax - Ymin)}$$

$$+ \frac{(Z - Zmin) (COS(Tilt))}{(Zmax - Zmin)}$$

where Xmin, Xmax, Ymin, Ymax, Zmin, and Zmax are the minimum and maximum values of the axes. Rotation and Tilt are the angles specified for the tilt and rotation of the axes.

Andrew's Plot

Object of Program

This plot takes multidimensional data and plots it on a two-dimensional plotting device in a meaningful way. It does this by mapping the vector $X = (X_1, X_2, X_3, \ldots, X_k)$ into a function of the form $Fx(t) = X_1 \sqrt{2} + X_2 \sin(t) + X_3 \cos(t) + X_4 \sin(2t) + X_5 \cos(2t) + \ldots$ where t is between $\pm\pi$. For further information, see Reference 1.

Special Considerations

1. Up to twenty variables may be used for plotting.
2. Each observation causes one line to be plotted.
3. The order of the variables determines the outcome of the plot.
4. Neither axis may be labeled.
5. A rough guess is made by the program as to extremes of the functions being plotted, and may be modified by pressing the "YAXIS" special function key.
6. The duration of the plot increases with the number of variables being used.
7. Any time a missing value is encountered for any variable used, the entire observation is deleted. For labeling the lines, the observation number that would have been used to label the line is incremented and used for the next observation.
8. Each line being plotted is broken up into 100 straight-line increments.

Special Plotting Characteristics

PLOT

This option creates the Andrew's Plot. You may choose whether or not you wish to have the first twenty observations labeled. Because this plot constructs one curve for every observation, it takes quite awhile to complete the plot for large data sets.

X-AXIS

In changing the parameters for the x-axis, the minimum value of x must be between (−PI) and (+PI). The maximum value must be between the minimum value and (+PI).

Labels

The only label that may be changed is the title.

References

1. D. F. Andrews, "Plots of High-Dimensional Data", Biometrics, 28, pp. 125-136, March, 1972.

Examples

STATISTICAL GRAPHICS EXAMPLES

```
***************************************************************************
*                        DATA MANIPULATION                              *
***************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                  Raw data to be input
Mode number = ?
2                                                  From mass storage
Is data stored on program's scratch file (DATA)?
YES
```

```
                    EGG FUTURE CONTRACTS

Data file name: DATA

Data type is:   Raw data

Number of observations:    83
Number of variables:        5

Variable names:
   1. ALBUMEN
   2. FROZ. ALBU
   3. FROZ. EGGS                                   Five variables and names or labels
   4. SHELLEGGS
   5. EGG.FUTURE

Subfile name     beginning observation  number of observations
 1. SUBFILE 1                        1                      30
 2. SUBFILE 2                       31                      12
 3. SUBFILE 3                       43                      24
 4. SUBFILE 4                       67                      17

SELECT ANY KEY
                                                   Press special function key labeled-LIST
Option number = ?
1                                                  All data listed
Enter method for listing data:
3
```

EGG FUTURE CONTRACTS

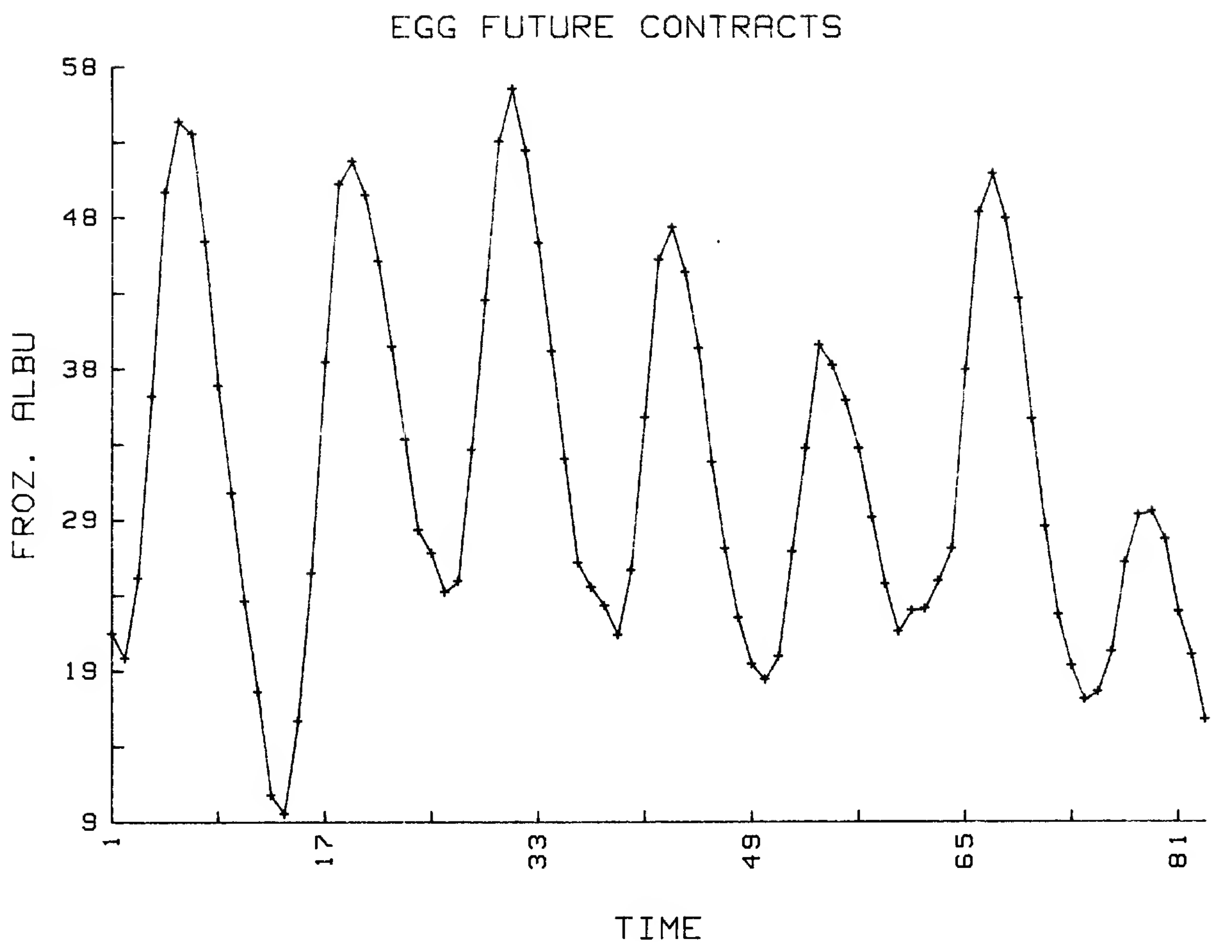
Data type is. Raw data

OBS#	Variable # 1 (ALBUMEN)	Variable # 2 (FROZ. ALBU)	Variable # 3 (FROZ. EGGS)	Variable # 4 (SHELLEGGS)	Variable # 5 (EGG.FUTURE)
1	1.67000	21.20000	2103.00000	.20000	43.58000
2	1.80000	19.60000	2025.00000	.20000	47.90000
3	1.99000	24.80000	2834.00000	.30000	47.40000
4	1.92000	36.60000	4697.00000	.50000	45.10000
5	1.92000	49.80000	6842.00000	1.20000	43.00000
6	2.12000	54.40000	7793.00000	2.10000	42.85000
7	2.34000	53.60000	7920.00000	2.30000	42.15000
8	2.38000	46.60000	6979.00000	2.20000	40.85000
9	2.26000	37.30000	5740.00000	1.70000	41.75000

10	2.08000	30.30000	4627.00000	1.10000	43.10000
11	2.06000	23.30000	3392.00000	.80000	43.00000
12	2.02000	17.40000	2429.00000	.30000	46.90000
13	1.96000	10.70000	1912.00000	.10000	46.45000
14	1.81000	9.50000	1681.00000	.30000	45.15000
15	1.83000	15.50000	2179.00000	.30000	44.70000
16	1.61000	25.10000	3425.00000	.30000	44.50000
17	1.53000	38.80000	5294.00000	.60000	45.40000
18	1.55000	50.30000	6464.00000	1.20000	42.80000
19	1.42000	51.80000	6431.00000	1.50000	41.00000
20	1.36000	49.60000	5955.00000	1.30000	37.00000
21	1.25000	45.30000	5186.00000	1.00000	37.00000
22	1.23000	39.80000	4478.00000	.70000	39.50000
23	1.19000	33.80000	3734.00000	.60000	39.75000
24	1.18000	27.90000	2930.00000	.50000	40.60000
25	1.15000	26.40000	2599.00000	.30000	39.90000
26	1.16000	23.90000	2527.00000	.30000	40.20000
27	1.20000	24.60000	3304.00000	.50000	37.55000
28	1.28000	33.10000	4388.00000	.90000	36.60000
29	1.45000	42.80000	5907.00000	1.20000	36.50000
30	1.55000	53.10000	6836.00000	1.70000	34.05000
31	1.33000	56.50000	6769.00000	1.80000	35.70000
32	1.20000	52.50000	6074.00000	1:50000	35.00000
33	1.17000	46.50000	5148.00000	1.20000	34.58000
34	1.22000	39.50000	4101.00000	.90000	41.25000
35	1.16000	32.50000	3174.00000	.60000	43.30000
36	1.05000	25.80000	2329.00000	.30000	43.10000
37	1.03000	24.20000	1921.00000	.20000	41.65000
38	1.00000	23.00000	1749.00000	.20000	41.70000
39	1.06000	21.10000	1535.00000	.10000	42.50000
40	1.07000	25.30000	2176.00000	.10000	43.10000
41	1.10000	35.20000	3437.00000	.30000	41.05000
42	1.09000	45.40000	4448.00000	.70000	39.95000
43	.96000	47.50000	4459.00000	.90000	40.15000
44	.91000	44.60000	4103.00000	.70000	37.65000
45	.87000	39.70000	3423.00000	.50000	41.75000
46	.80000	32.30000	2711.00000	.30000	37.80000
47	.80000	26.70000	2112.00000	.20000	36.80000
48	.84000	22.20000	1631.00000	.10000	36.00000
49	.88000	19.20000	1249.00000	.10000	36.50000
50	.84000	18.20000	1209.00000	.10000	36.70000
51	.83000	19.70000	1500.00000	.10000	35.70000
52	.83000	26.50000	2687.00000	.10000	32.70000
53	.81000	33.20000	4024.00000	.50000	31.50000
54	.81000	39.90000	4831.00000	1.00000	32.40000
55	.81000	38.60000	4739.00000	1.10000	31.25000
56	.81000	36.30000	4513.00000	.90000	28.30000
57	.70000	33.20000	3966.00000	.70000	29.00000
58	.74000	28.70000	3489.00000	.60000	35.35000
59	.84000	24.40000	2732.00000	.50000	34.95000
60	.75000	21.30000	2180.00000	.30000	36.60000
61	.73000	22.70000	2210.00000	.20000	35.80000
62	.67000	22.80000	2322.00000	.30000	34.10000
63	.68000	24.60000	2243.00000	.30000	36.00000
64	.85000	26.70000	2580.00000	.20000	37.85000
65	.85000	38.30000	3836.00000	.30000	38.60000
66	.88000	48.50000	5086.00000	.80000	35.70000
67	.88000	51.00000	5241.00000	1.10000	34.95000
68	.81000	48.10000	4748.00000	1.00000	34.65000
69	.75000	42.90000	4022.00000	.70000	35.45000
70	.69000	35.10000	3149.00000	.50000	38.50000
71	.68000	28.10000	2307.00000	.30000	37.00000
72	.71000	22.40000	1700.00000	.10000	36.35000
73	.75000	19.10000	1456.00000	.10000	38.15000
74	.76000	16.90000	1282.00000	.10000	38.70000
75	.85000	17.40000	1417.00000	.20000	36.35000
76	.84000	20.00000	1772.00000	.20000	37.00000

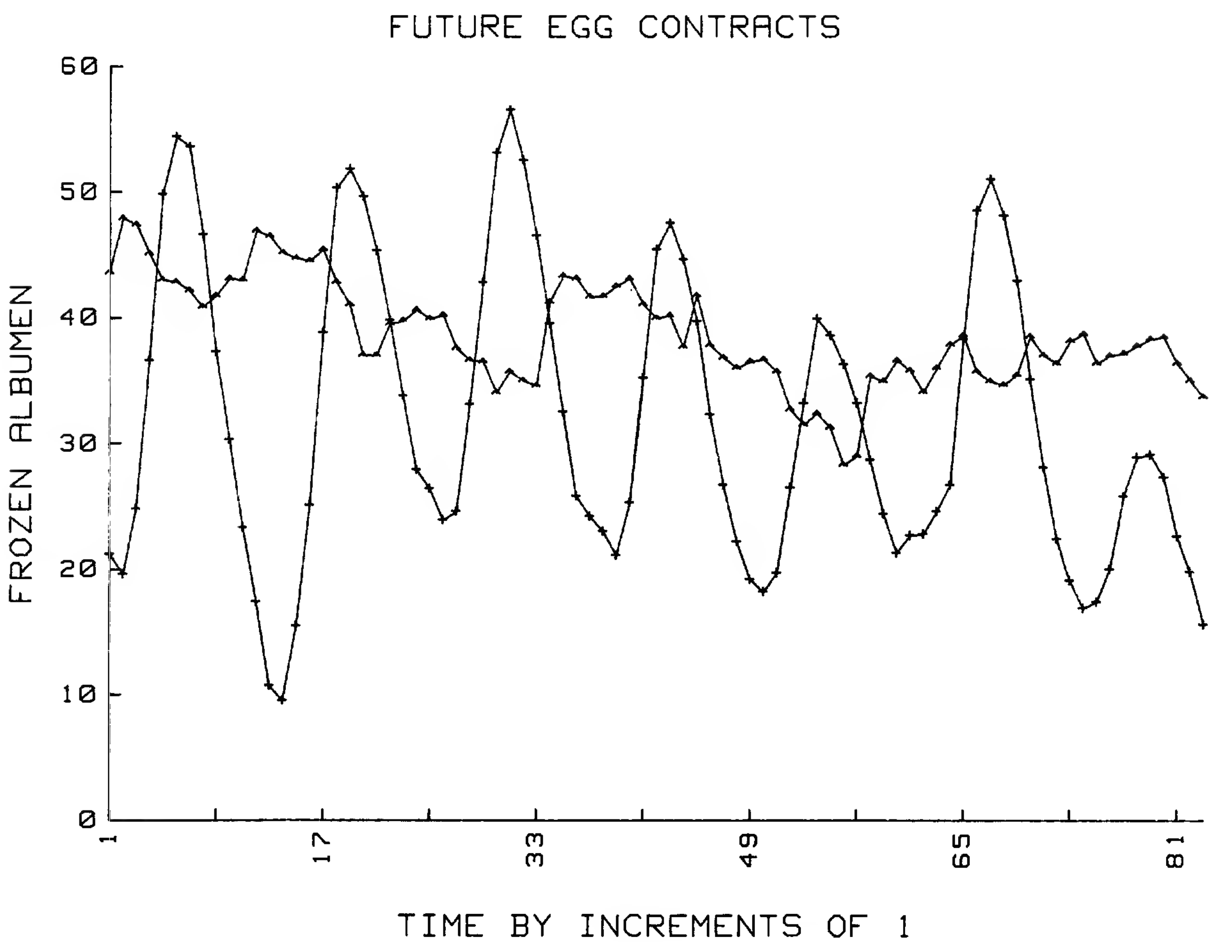
77	.95000	25.80000	2578.00000	.10000	37.15000
78	.98000	28.90000	3215.00000	.20000	37.75000
79	.98000	29.10000	3165.00000	.40000	38.30000
80	1.05000	27.30000	3025.00000	.30000	38.45000
81	1.00000	22.60000	2746.00000	.30000	36.35000
82	.90000	19.80000	2311.00000	.20000	35.00000
83	.92000	15.60000	1853.00000	.10000	33.70000

Option number = ?	
0	Exit listing options
SELECT ANY KEY	Press special function key labeled-ADV STATS
	Remove BSDM media
	Insert Statistical Graphics 1A media
Enter number of desired function:	
1	Time Plot
Y axis variable number?	
2	
Enter subfile to be used (0 if subfiles ignored)	
0	
Enter number of desired function:	
8	Select plotter option
Enter option number of the graphics device?	
2	Choose external plotter
Plotter identifier string (press CONT if 'HPGL')?	Press CONTINUE
Enter select code, bus address (default is 7,5) ?	Press CONTINUE
Is the above information correct?	
YES	
Enter number of desired function:	
1	Plot
Are the points to be connected?	
YES	
Are grid lines to be plotted?	
NO	
Beep will sound when plot is done then press CONT.	
To interrupt plotting press STOP key.	




```
Enter number of desired function:
4                                                  Change y-axis
Y plotting minimum?
0
Y plotting maximum?
60
Y tic ?
10
Label every Kth tic mark?
1
Number of decimal places to label the Y axis?
0
Enter number of desired function:
5                                                  Change labels
Enter the Time axis title (33 characters or less)
TIME BY INCREMENTS OF 1
Enter the Y axis title (33 characters or less)
FROZEN ALBUMEN
Enter the Graph Title (33 characters or less)
FUTURE EGG CONTRACTS
Enter number of desired function:
1                                                  Plot
Are the points to be connected?
YES
Are grid lines to be plotted?
NO
```

```
Beep will sound when plot is done then press CONT.
To interrupt plotting press STOP key.

Enter number of desired function:
10                                                    Overlay plot
Y axis variable number?
5
Enter subfile to be used (0 if subfiles ignored)
0
Enter number of desired function:
6                                                     Change plotting character
Put double quotes around the blank.
^
Enter number of desired function:
1                                                     Plot
Are the points to be connected?
YES
Beep will sound when plot is done then press CONT.
To interrupt plotting press STOP key.
```



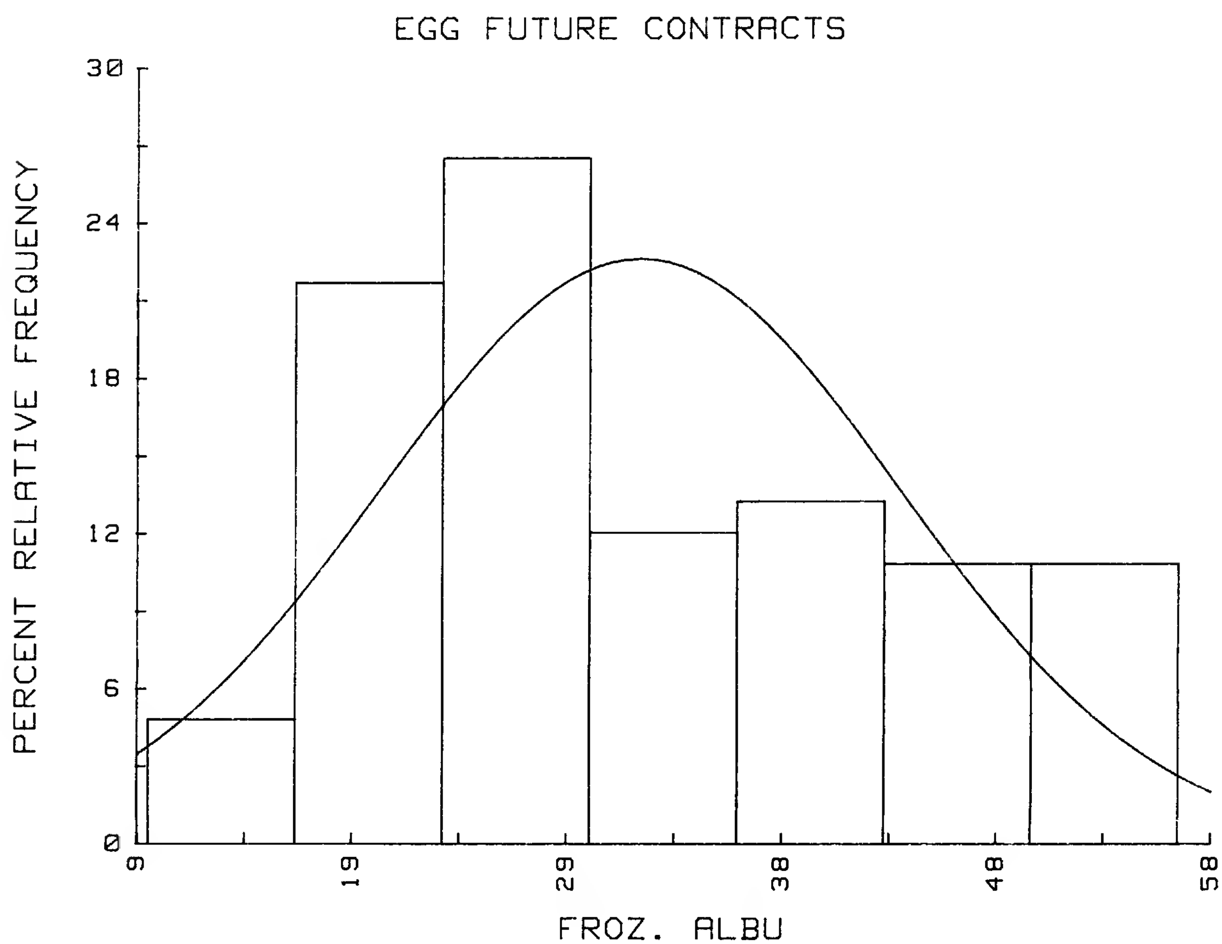

```
Enter number of desired function:
11                                                    Store plotting characteristics
Enter file name to store plot characteristics ?
CHARS:INTERNAL
```

```
Is data medium placed in device INTERNAL
?
YES
Is PROGRAM MEDIUM replaced in device
?
?
YES
Enter number of desired function:
13                                                      Return to main graphics menu.

Enter number of desired function:
2                                                       Select histogram example


HISTOGRAM


Variable number for creating histogram?
2

Variable 2 will be used

Enter subfile to be used (0 if subfiles ignored)
0
Number of valid cases = 83
The mean is calculated to be= 31.9313253012
The variance is calculated to be= 140.299006759


                                       OBSERVED
CELL     MINIMUM     MAXIMUM     FREQUENCY
  1        9.500      16.214         4
  2       16.214      22.929        18
  3       22.929      29.643        22
  4       29.643      36.357        10
  5       36.357      43.071        11
  6       43.071      49.786         9
  7       49.786      56.500         9

Enter number of desired function:
8                                                       Select plotter
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)

Is the above information correct?
YES
Enter number of desired function:
1                                                       Plot
Are horizontal grid lines to be plotted?
NO
BEEP will sound when plot done then Press CONT.
To interupt plotting, press STOP key.


Enter number of desired function:
10                                                      Overlay normal curve
```

CELL	MINIMUM	MAXIMUM	OBSERVED FREQUENCY	EXPECTED FREQUENCY	CONTRIBUTION TO CHI-SQUARE
1	-Infinity	16.214	4	7.658	1.748
2	16.214	22.929	18	10.901	4.623
3	22.929	29.643	22	16.583	1.770
4	29.643	36.357	10	18.448	3.869
5	36.357	43.071	11	15.011	1.072
6	43.071	49.786	9	8.932	.001
7	49.786	Infinity	9	5.466	2.284

Press CONT to plot the normal curve overlay

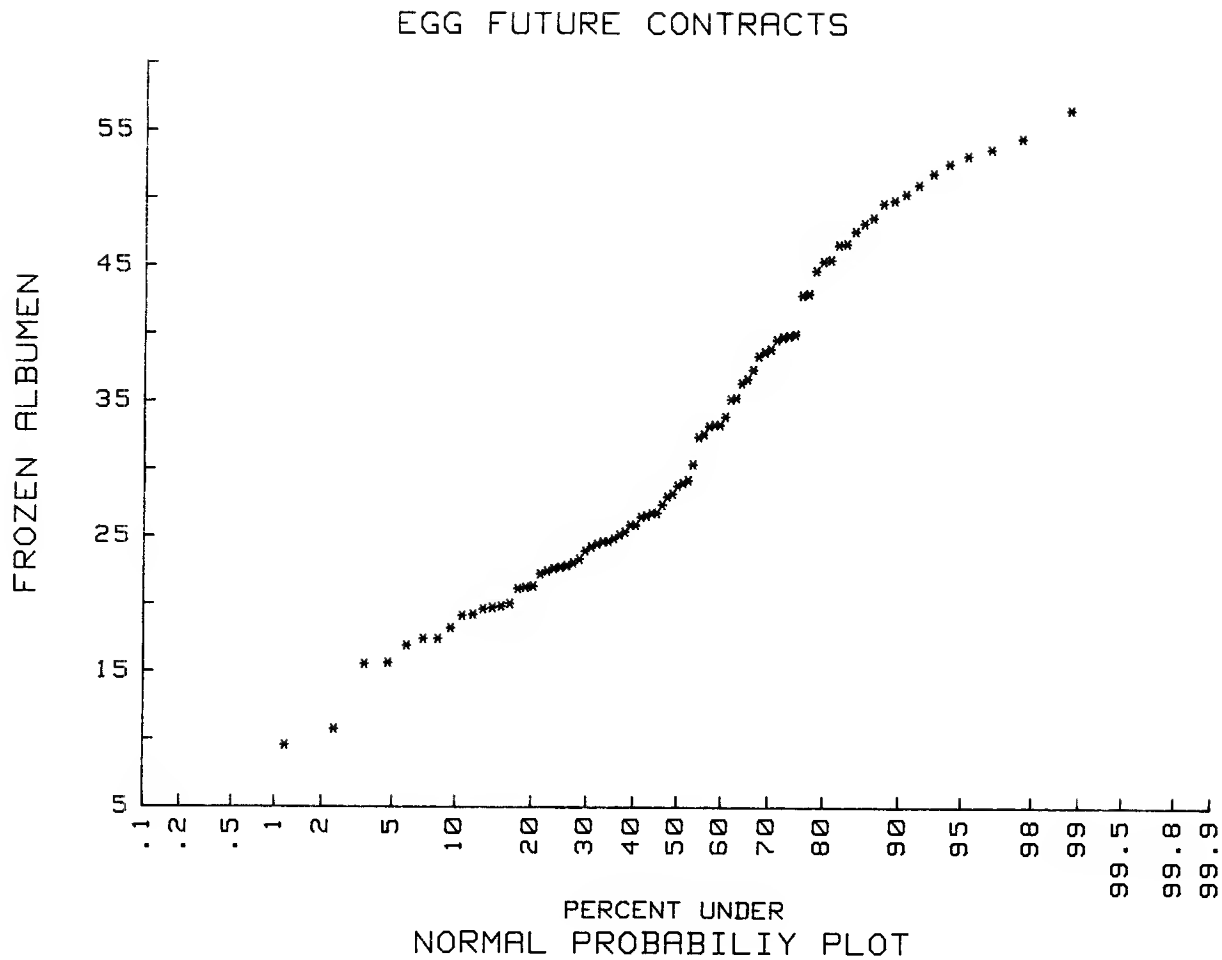
BEEP will sound when plot done then PRESS CONT.



Enter number of desired function:
13 — Return to main graphics menu

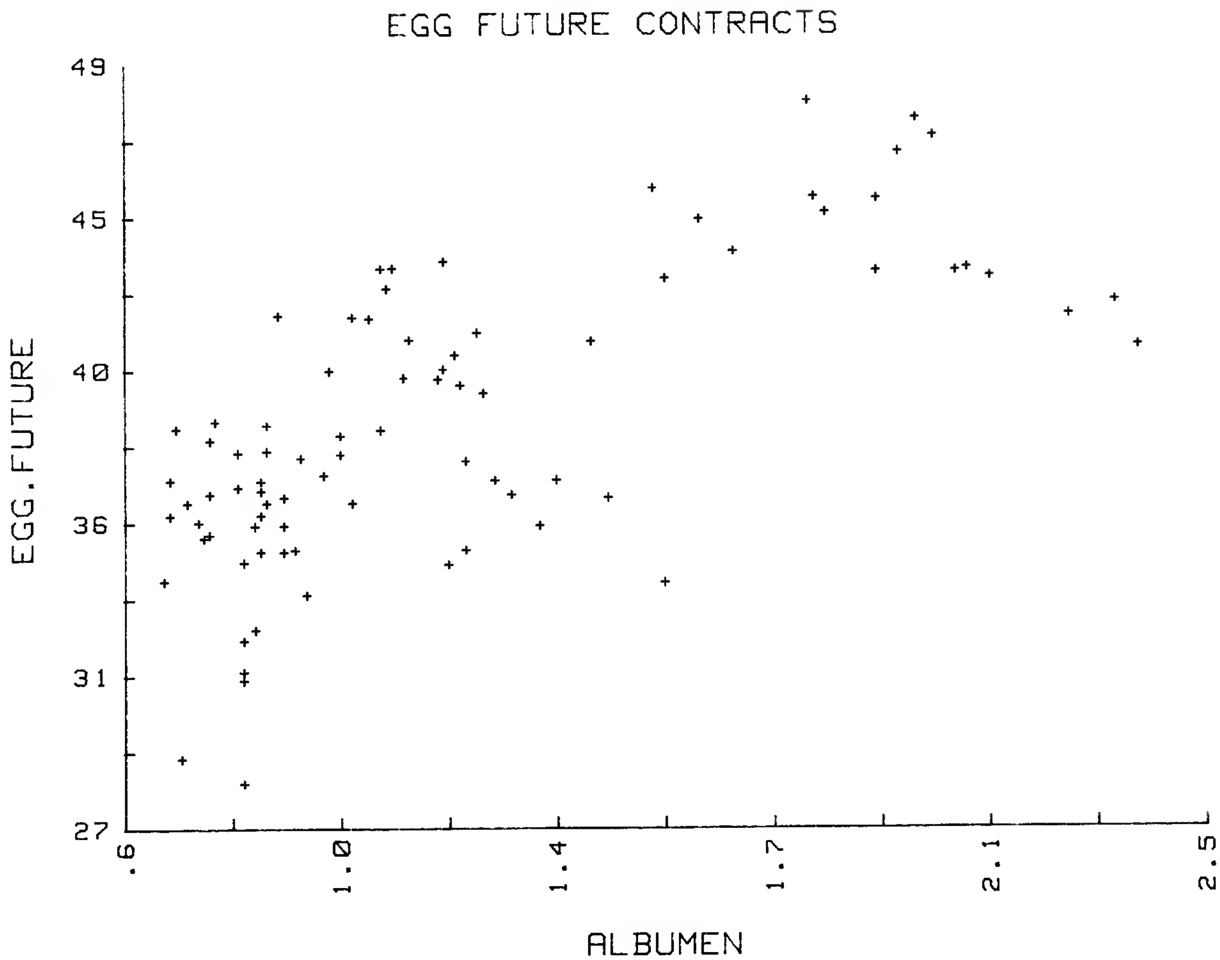
Enter number of desired function:
3 — Select normal probability plot
Variable number?
2
Enter subfile to be used (0 if subfiles ignored)
0
SORTING THE DATA
Enter number of desired function:
3 — Change y-axis

```
Y plotting minimum?
5                                                      Specify y lower limit
Y plotting maximum?
60                                                     Specify y upper limit
Y tic ?
5
Label every Kth tic mark?
1                                                      Label every tic mark
Number of decimal places for labeling the Y axis?
0
Enter number of desired function:
4                                                      Change labels and titles
Enter the Y axis title (33 characters or less)
FROZEN ALBUMEN
Enter the Graph Title (33 characters or less)
EGG FUTURE CONTRACTS
Enter number of desired function:
7                                                      Select plotter
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)

Is the above information correct?
YES
Enter number of desired function:
5                                                      Change plotting symbol
Put double quotes around the blank?
*
Enter number of desired function:
1                                                      Plot
Are grid lines to be plotted?
NO
Beep will sound when the plot done then press CONT
To interrupt plotting, press STOP key.
```




```
Enter number of desired function:
12                                                    Return to main graphics menu

Enter number of desired function:
5                                                     Select scattergram
X axis variable number?
1
Y axis variable number?
5
Enter subfile to be used (0 if subfiles ignored)
0
Enter number of desired function:
8                                                     Select plotter option
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)?

Is the above information correct?
YES
Enter number of desired function:
1                                                     Plot
Are the points to be connected?
NO
```

```
Are grid lines to be plotted ?
NO
Beep will sound when plot done then press CONT.
To interrupt plotting press 'STOP' key.
```

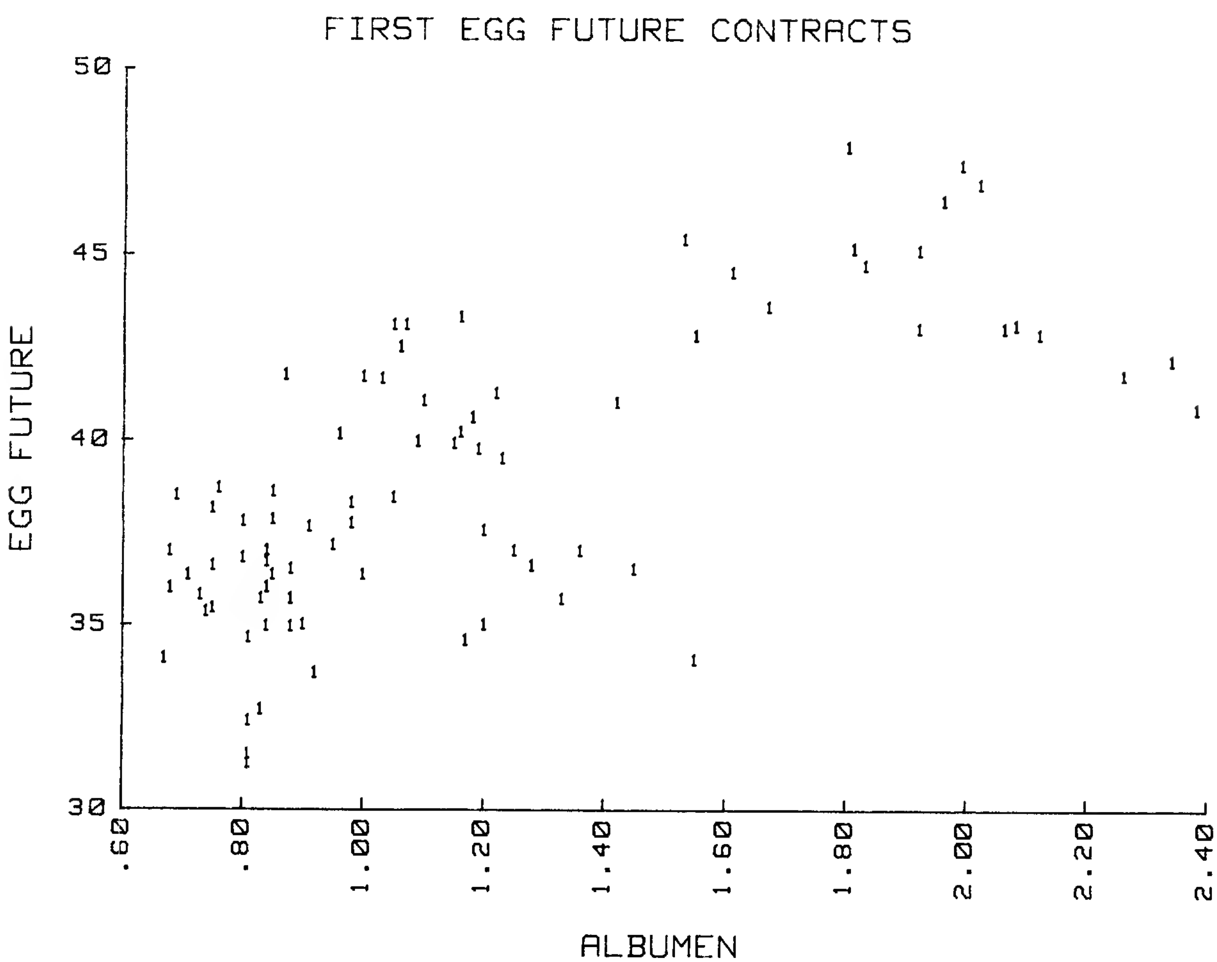



```
Enter number of desired function:
4
Y plotting minimum?
30
Y plotting maximum?
50
Y tic?
5
Label every Kth tic mark?
1
Number of decimal places for labeling the Y axis?
0
Enter number of desired function:
3
X plotting minimum?
.6
X plotting maximum?
2.4
X tic?
.2
```

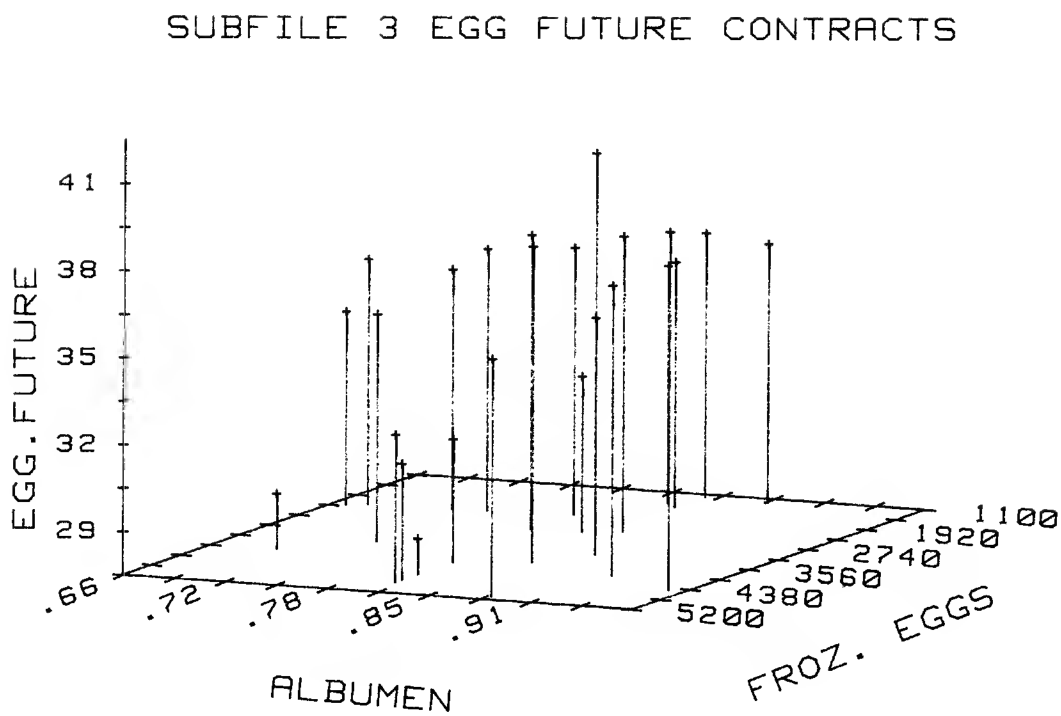
Change y-axis for another scattergram

Change x-axis

```
Label every Kth tic mark?
1
Number of decimal places for labeling the X axis?
2
Enter number of desired function:
6                                                      Change plotting symbol
Put double quotes around the blank?
1
Enter number of desired function:
5                                                      Change labels
Enter the X axis title (33 characters or less)
ALBUMEN
Enter the Y axis title (33 characters or less)
EGG FUTURE
Enter the Graph Title (33 characters or less)
FIRST EGG FUTURE CONTRACTS
Enter number of desired function:
1                                                      Plot
Are the points to be connected?
NO
Are grid lines to be plotted ?
NO
Beep will sound when plot done then press CONT.
To interrupt plotting press 'STOP' key.
```

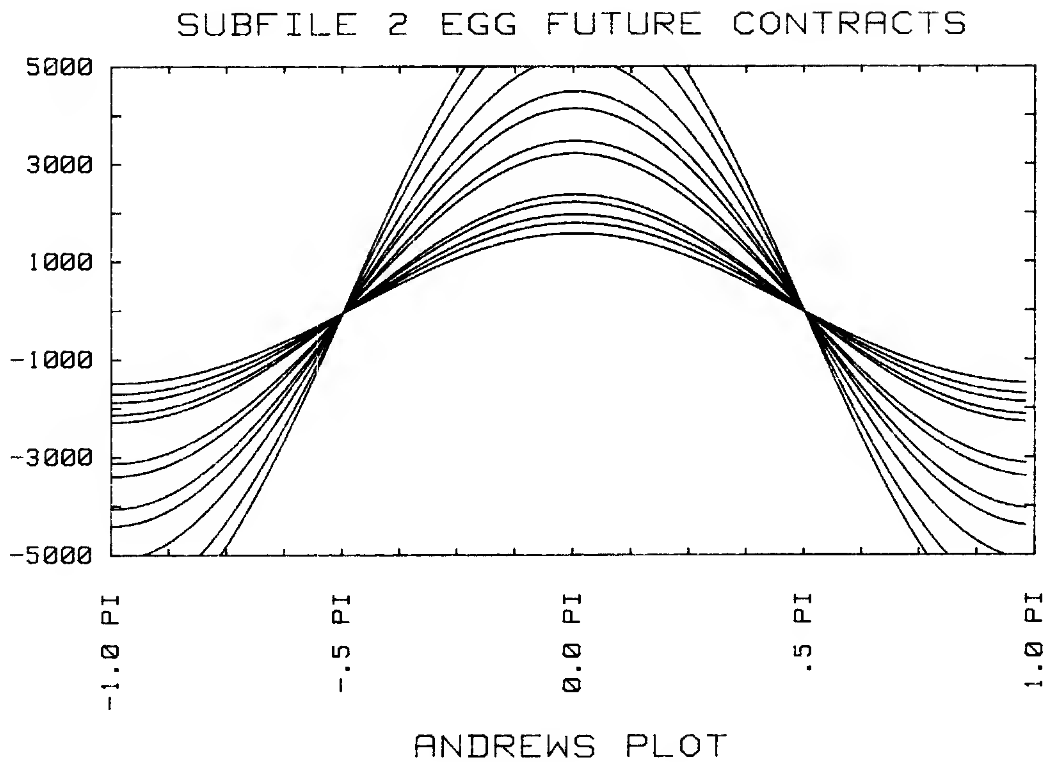


Enter number of desired function:	
13	Return to main graphics menu
Enter number of desired function:	
7	Select another ADV STAT pac
	Remove Statistical Graphics 1A
	Insert Statistical Graphics 1B
Enter number of desired function:	
3	Select 3-D plot
X axis variable number?	
1	
Y axis variable number?	
3	
Z axis variable number?	
5	
Enter subfile to be used (0 if subfiles ignored)	
3	Plot only for data in subfile 3.
Enter number of desired function:	
9	Select plotter
Enter option number of the graphics device	
2	
Plotter identifier string (press CONT if 'HPGL' ?	
Enter select code, bus address (defaults are 7,5)?	
IS THE ABOVE INFORMATION CORRECT?	
YES	
Enter number of desired function:	
1	Plot
Enter angle of rotation in degrees [0<Angle<=90]	
30	Rotate plot for easier viewing
Enter angle of elevation in degrees [0<=Angle<=90]	
30	Raise angle of elevation
Beep will sound when plot done then PRESS CONT.	
To interrupt plotting press 'STOP' key.	



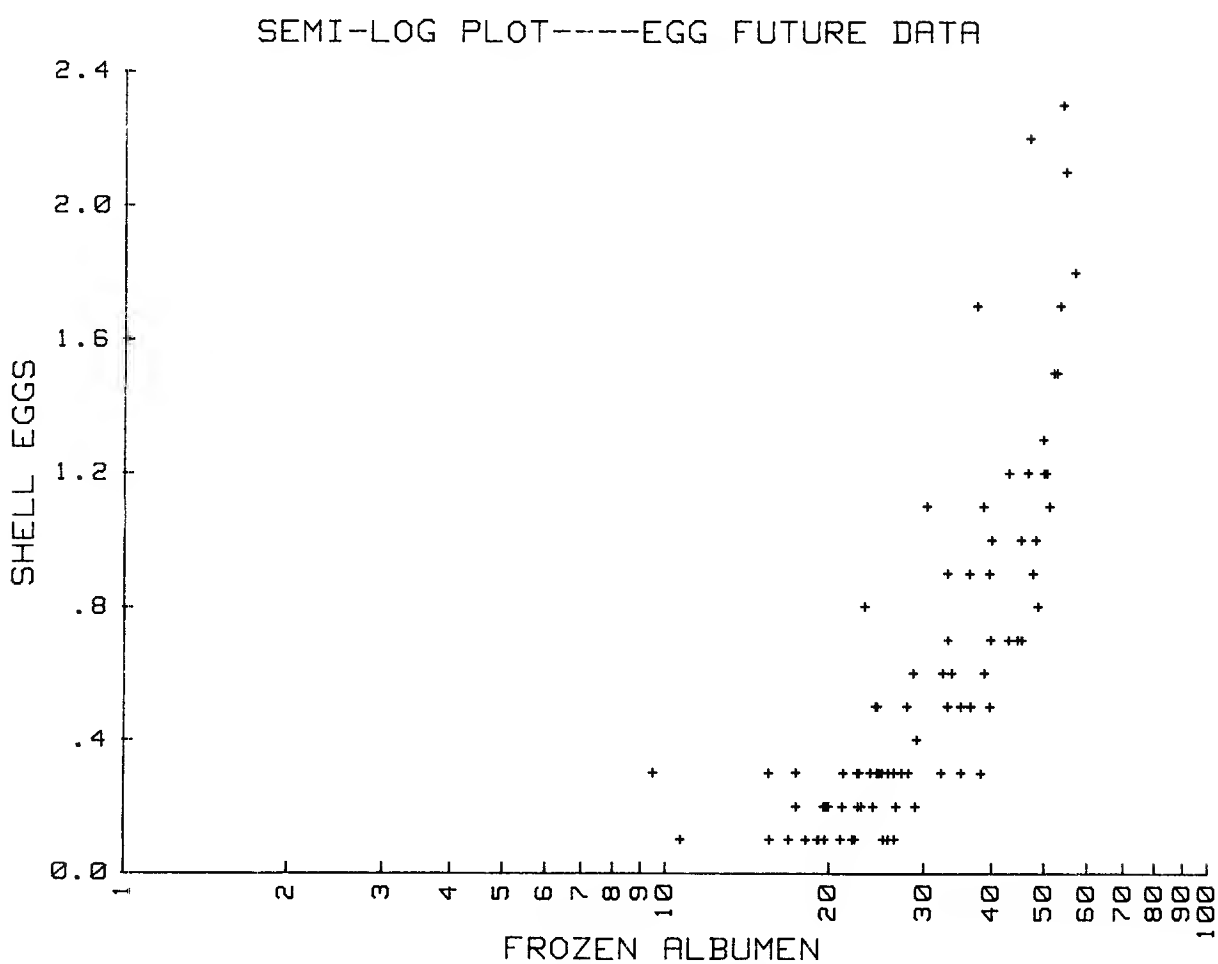

```
Enter number of desired function:
14                                                      Return to main graphics menu

Enter number of desired function:
4                                                       Select Andrews Plot
Number of variables to be used?
5
Enter variable number 1
?
1
Enter variable number 2
?
2
Enter variable number 3
?
3
Enter variable number 4
?
4
Enter variable number 5
?
5
Is the above information correct?
YES
Enter subfile to be used (0 if subfiles ignored)
2                                                       Plot only data in subfile 2
Enter number of desired function:
7                                                       Select plotter
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)?

Is the above information correct?
YES
```

```
Enter number of desired function:
1                                                    Plot
Are up to the first twenty lines to be labelled?
YES
Beep will sound when plot done then PRESS CONT.
To interupt the plot press the STOP key
```



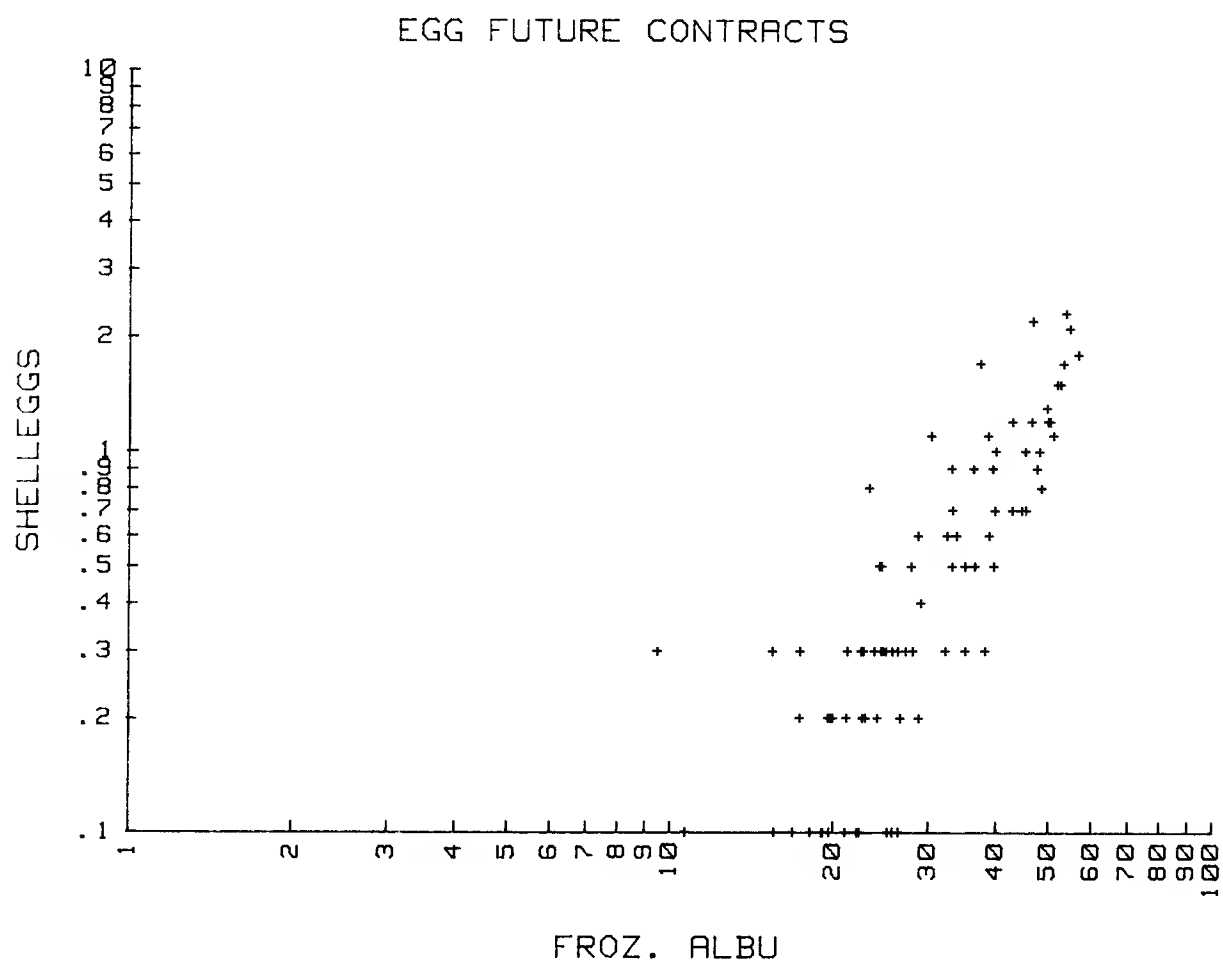

```
Enter number of desired function:
12                                                   Return to main graphics menu

Enter number of desired function:
1                                                    Select semi-log plot
X axis (LOG AXIS) variable number?
2
Y axis variable number?
4
Enter subfile to be used (0 if subfiles ignored)
0
Enter number of desired function:
3                                                    Change y-axis
Y plotting minimum?
0
Y plotting maximum?
2.4
Y tic?
.4
Label every Kth tic mark?
1
Number of decimal places for labeling the Y axis?
1
Enter number of desired function:
4                                                    Change labels
```

```
Enter the X axis title (33 characters or less)
FROZEN ALBUMEN
Enter the Y axis title (33 characters or less)
SHELL EGGS
Enter the Graph Title (33 characters or less)
SEMI-LOG PLOT----EGG FUTURE DATA
Enter number of desired function:
7                                                     Select plotter
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)?

Is the above information correct?
YES
Enter number of desired function:
1                                                     Plot
Are grid lines to be plotted?
NO
Beep will sound when plot is done then press CONT
To interrupt plotting, press 'STOP' key
```

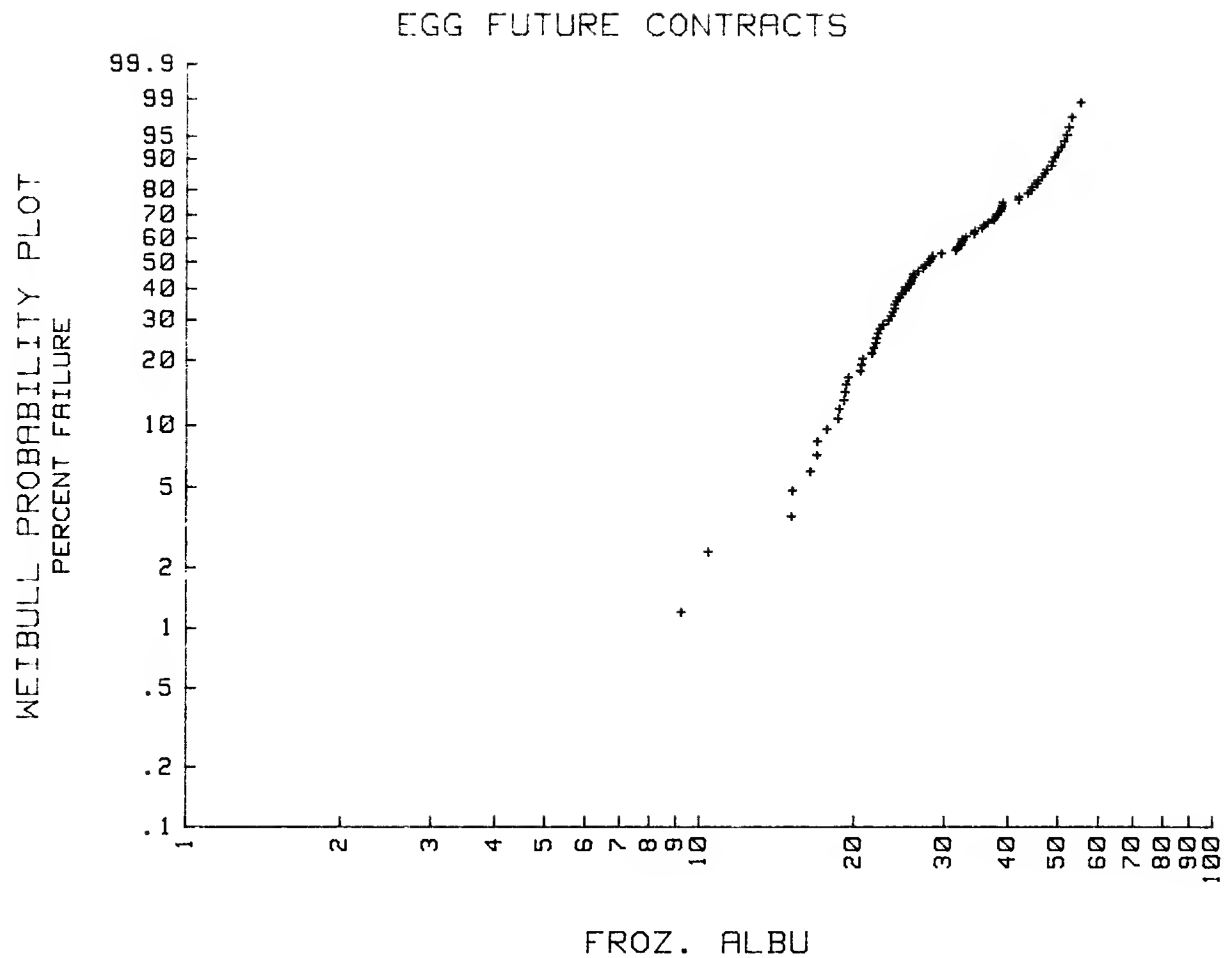



```
Enter number of desired function:
12                                                    Return to main graphics menu
```

```
Enter number of desired function:
2                                                     Select log-log plot
X axis variable number?
2
Y axis variable number?
4
Enter subfile to be used (0 if subfiles ignored)?
0
Enter number of desired function:
6                                                     Select plotter
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)?

Is the above information correct?
YES
Enter number of desired function:
1                                                     Plot
Are grid lines to be plotted?
NO
Beep will sound when plot done then press CONT.
To interrupt plotting, press 'STOP' key.
```



```
Enter number of desired function:
11                                                        Return to main graphics menu
Enter number of desired function:
6                                                         Return to statistical graphics 1A


Enter number of desired function:
4                                                         Select Weibull Plot
Variable number?
2
Enter subfile to be used (0 if subfiles ignored)
0
SORTING THE DATA
Enter number of desired function:
6                                                         Select plotter
Enter option number of the graphics device?
2
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, bus address (default is 7,5)?

Is the above infromation correct?
YES
Enter number of desired function:
1                                                         Plot
Are grid lines to be plotted?
NO
Beep will sound when plot done then press CONT.
To interrupt plotting, press 'STOP' key.
```




```
Enter number of desired function:
11
```

Return to main graphics menu

```
Enter number of desired function:
6
```

Return to Basic Statistics and Data Manipulation (BSDM)

General Statistics

General Information

Description

The General Statistics module includes 5 major parts:

1. **One Sample Tests** allow you to run a series of tests and plots on one-variable problems. You can test whether the observations are mutually independent, whether the mean of the data is significantly different from a hypothesized mean, compare your data with normal, exponential, or uniform distributions, and test the randomness of your data.
2. **Paired-Sample Tests** allow you to compare the means of two samples, test if the paired samples are similar, fit the data with a regression equation, test whether the two populations have the same median and test the independence of two random variables.
3. **Two-Independent-Sample Tests** allow you to test whether the means of two samples are equal, whether the medians of two samples are equal, and whether the two populations have the same distribution.
4. **Multiple-Sample (≥3 Samples) Tests** allow you to test whether the means of several populations are equal, and whether there are significant differences between pairs of means.
5. **Statistical Distributions** allow you to study a series of continuous and discrete statistical distributions. Both tabled values and right-tailed probabilities are available for the continuous distributions. The discrete distributions calculate right-tail probabilities, single term probabilities and an approximate value for a specified right-tailed probability. This program will also calculate n factorial, the complete gamma function, the complete beta function and binomial coefficients.

Methods and Formulae, References, etc., for each of these five parts are found in each of the following sections.

Special Considerations

If you specify one type of test (for example, Paired-Sample Tests), you will not be able to perform a different type of test (say, Multiple-Sample Tests), without returning to the Start-up procedure for the new test. You must access the Start-up procedure to define the segment of the data matrix which is to be tested.

One Sample Tests

Object of Programs

This section allows you to run a series of tests and plots on one variable (or one subfile of one variable) from the data matrix defined by the Basic Statistics and Data Manipulation program. Each test will automatically sort or restore the data to its original form as needed. You can perform several kinds of tests on your data:

Serial Correlation — tests if the observations are mutually independent.

t-Test — tests if the mean of the data is significantly different from a hypothesized mean which you specify.

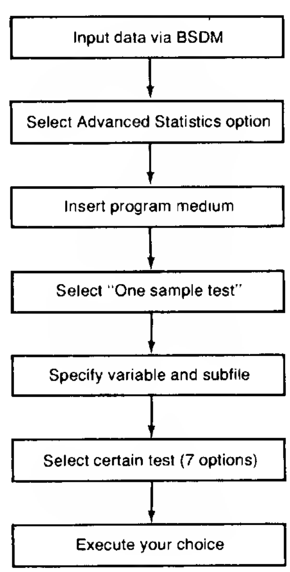
Kolmogorov-Smirnov Goodness-of-fit test or Chi-Square Goodness-of-fit — test if your data follow a normal, exponential or uniform distribution.

Runs Test — tests the randomness of your data.

Shapiro-Wilk Test — tests for normality.

The above tests will be described in Methods and Formulae.

Typical Program Flow



Data Structure

Since we have only one variable, the data is entered as in the following example, which shows a sample of size 12:

Variable #1

I	OBS(I)	OBS(I+1)	OBS(I+2)	OBS(I+3)	OBS(I+4)
1	2	5	8	7	3
6	6	4	5	9	7
11	3	4			

Alternatively, you may input a data set containing several variables, then specify a **single** variable for the analysis. Several variables may be analyzed in succession.

Methods and Formulae

Basic Statistics

For the calculation of the sample mean, variance, standard deviation, standard error of the mean, coefficient of variation, skewness, kurtosis, and confidence intervals on the mean and variance, please refer to Snedecor and Cochran's Statistical Methods.

Kolmogorov-Smirnov Goodness-of-Fit Test

- Assumptions

1. The sample is a random sample.
2. If the hypothesized distribution function G(X), in H0 below, is continuous the test is exact. Otherwise, the test is conservative.

- Hypotheses

Let G(X) be a completely specified, hypothesized distribution function. F(X) is the distribution function for the random variable X.

1. Two-Sided Test
   H0: $F(X) = G(X)$ for all X.
   H1: $F(X) \neq G(X)$ for at least one value of X.
2. One-Sided Test
   H0: $F(X) \geq G(X)$ for all X.
   H1: $F(X) < G(X)$ for at least one value of X.
3. One-Sided Test
   H0: $F(X) \leq G(X)$ for all X.
   H1: $F(X) > G(X)$ for at least one value of X.

- Test Statistics

Let $S(X)$ be the empirical distribution function based on the random sample $X1, X2, \ldots, Xn$.

1. Two-Sided Test

   Let the test statistic T be the greatest (denoted by "sup" for supremum) vertical distance between $S(X)$ and $G(X)$.

   $T = \sup|G(X)\text{-}S(X)|$

2. One-Sided Test
   $T1 = \sup\,[G(X)\text{-}S(X)]$

3. One-Sided Test
   $T2 = \sup[S(X)\text{-}G(X)]$

- Decision Rule

Reject H0 at the level of significance α if the appropriate test statistic, T, T1, or T2 exceeds the $1-\alpha$ quantile $W(1-\alpha)$ from the Table of Quantiles of the Kolmogorov Test Statistic.

Chi-square Goodness-of-Fit Test

- Assumptions

1. The sample is a random sample.
2. The measurement scale is at least nominal.

- Hypothesis

Let $F(X)$ be the true but unknown distribution function and let $G(X)$ be a completely specified, hypothesized distribution function.

H0: $F(X) = G(X)$ for all X.
H1: $F(X) \neq G(X)$ for at least one X.

- Test Statistic

Suppose the data is divided into c classes, and the number of observations falling in each class is denoted by Oj, for $j = 1, 2, \ldots, c$. Let Pj be the probability of a random observation being in class j under the assumption that $G(X)$ is the distribution function of X. Then define Ej as $Ej = Pj*n$, where n is the sample size. Then, the test statistics is:

$$T = \Sigma\,(Oj - Ej)^2/Ej \qquad \text{for } j = 1, 2, \ldots, c.$$

- Decision Rule

The exact distribution of T is difficult to use, so the large sample approximation is used. The approximate distribution of T is the Chi-square distribution with $(c-1)$ degrees of freedom. Therefore, the critical region of approximate size α corresponds to values of T greater than $\chi^2(1-\alpha)$, the $(1-\alpha)$ quantile of a χ^2 random variable with $(c-1)$ degrees of freedom. Reject H0 if T exceeds $\chi^2(1-\alpha)$; otherwise, accept H0.

t-Test
Let $X1, \ldots, Xn$ be a random sample from a population with mean μ, where M is the sample mean and S is the sample standard deviation.

- Hypotheses
  1. Two-Sided
     H0: $\mu = a$, the hypothesized value for the population mean.
     H1: $\mu \neq a$
  2. One-Sided
     H0: $\mu = a$
     H1: $\mu < a$
  3. One-Sided
     H0: $\mu = a$
     H1: $\mu > a$

- Test Statistic

  $t = \sqrt{n}(M\text{-}a)/S$

- Decision Rule

The statistic t has a t-distribution with $(n-1)$ degrees of freedom. $T(1-\alpha, n-1)$ is the $(1-\alpha)$ quantile of the t-distribution with $(n-1)$ degrees of freedom.

1. Two-Sided: if $t \leqslant T(1-\alpha/2, n-1)$, accept H0, otherwise, reject H0.
2. One-Sided: if $t \geqslant T(\alpha/2, n-1)$, accept H0, otherwise, reject H0.
3. One-Sided: if $t \leqslant T(1-\alpha/2, n\text{-}1)$ accept H0, otherwise, reject H0.

In this program the corresponding one- or two-tailed probability of the computed t-value will be printed.

Runs Test
Any sequence of like observations bounded by observations of a different type is called a run. The number of observations in the run is called the length of the run.

Suppose a coin is tossed twenty times and the resulting heads (H) or tails (T) are recorded in the order in which they occur:

T HHHHHH T H T H TT HHH T H T H

Each segment is called a run. The total number of runs in the example is 12.

The total number of runs may be used as a measure of the randomness of the sequence; too many runs may indicate that each observation tends to follow and be followed by an observation of the other type, while too few runs might indicate a tendency for like observations to follow like observations. In either case the sequence would indicate that the process generating the sequence was not random.

- Hypothesis

H0: The process which generates the sequence is a random process.

H1: The random variables in the sequence are either dependent on other random variables in the sequence or are distributed differently from one another.

• Test Statistic

In this program we use the median as an indicator of two types of observations, i.e., a value below the median is one kind, a value above the median is another kind. Count the runs below and above the median, say D. Then

$$W = N + 1 + Z_p ([(N \uparrow 2)/(2N\text{-}1)] \uparrow .5)$$

where Z_p is the pth quantile of a standard normal random variable.

• Decision Rule

Reject H0 at the level α if $D > W(1-\alpha/2)$ or $D < W(\alpha/2)$, otherwise accept H0.

Serial Correlation
This routine checks for randomness in the sample.

• Formula

Serial correlation with lag k:

$$\left[\sum_{i=1}^{N-k} (X_i - \bar{X})(X_{i+k} - \bar{X}) \right] \Big/ \left[\sum_{i=1}^{N} X_i^2 - N \bullet \bar{X}^2 \right]$$

If the correlation is small, this means the observations are mutually independent.

Shapiro-Wilk Test
This routine performs a test for normality for a sample of size 3 to 50, inclusive.

Note

A tie means two or more observations have the same value. Ties must be given a special treatment when we try to give every single observation a rank.

If the sample size is less than 3 or greater than 50, a message will be printed stating that this program will not work and to try a chi-square goodness of fit test for N>50. Then you will have a chance to choose the test you want again.

• Hypothesis

The data comes from a normal distribution.

• Test Statistic

A test statistic W is printed followed by the tabled values of Wα (% POINTS) for alpha = .01, .02, .05, .1, and .5.

• Decision Rule

The observed test statistic W indicates that the sample did not come from a normal distribution at the corresponding alpha level of significance if the value of W is less than the corresponding percentage point. Hence, small values of W are significant.

References

1. Abramowitz, Milton and Stegun, Irene A (1970) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. U.S. Government Printing Office, Washington D.C., p. 949.
2. Box, G.E.P. and Cox, D.R. (1964). "An Analysis of Transformations". Journal of the Royal Statistical Society 26:2, pp. 211-252.
3. Conover, W.J. (1971). Practical Nonparametric Statistics. John Wiley & Sons, Inc., New York, p. 414.
4. Conte, S.D. (1965). Elementary Numerical Analysis. McGraw-Hill Book Company, New York, p. 135.
5. Dickinson Gibbons, Jean (1971). Nonparametric Statistical Inference. McGraw-Hill Book Company, New York, pp. 75-83.
6. Hahn, G. and Shapiro, S.S., (1967). Statistical Models in Engineering, John Wiley & Sons, Inc., New York, pp. 330-332.
7. Kopitzke, Robert W., Unpublished Notes, 1973.
8. Mood, Graybill, Boes (1974). Introduction to the Theory of Statistics, 3rd Edition, McGraw-Hill Book Company, New York. Chapter 7.
9. Shapiro, S.S. and Wilk, M.B. (1965). "An Analysis of Variance Test for Normality". Biometrika; 52, 3 and 4, p. 591.
10. Snedecor, George W. and Cochran, William G. (1967). Statistical Methods. The Iowa State University Press, Ames, Iowa.
11. Ullman, Neil R., (1972). Statistics: An Applied Approach, Xerox College Publishing, Lexington, Mass. pp. 354-357.

Paired-Sample Tests

Description

This program allows you to perform the following paired-sample tests:

Paired t-test — compare the means of two samples.

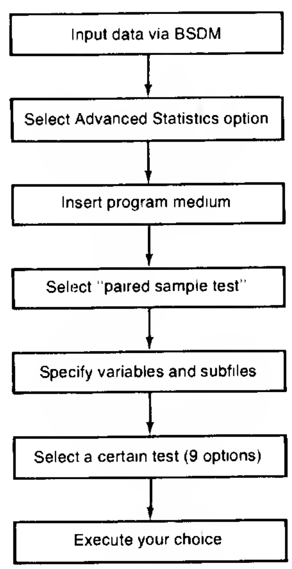
Cross Correlation — test if the paired samples are similar.

Family Regression — fit the data with one of several regression equations.

Sign Test or Wilcoxon Signed Rank Test — test whether two populations have the same median.

Spearman's Rho or Kendall's Tau — test the independence of two random variables.

Typical Program Flow



Data Structure

For paired-sample tests, two variables or the same subfile of two variables must be used.

The data are entered as in the following example:

Obs. #	Variable #1	Variable #2
1	54	46
2	44	42
3	46	44

Methods and Formulae

Paired t-Test

This is a one-sample t-test performed on the differences between paired samples. See the Methods and Formulae section in the One-Sample Tests chapter for further details.

Cross Correlaton

Provides a correlation between paired samples with a lag between them. Large values show the paired samples are quite similar, i.e., no significant difference. The cross correlation with lag k between the two samples X1,X2,...,XN and Y1,Y2,...YN is:

$$\left[\sum_{i=1}^{N-k} (X_i - \bar{X})(Y_{i+k} - \bar{Y}) \right] \Big/ \left[\sum_{i=1}^{N} (X_i - \bar{X})^2 \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \right] \uparrow .5$$

Family Regression

Provides four different regression models. All of the models are solved (except quadratic) by "linearizing" the model to the form:

$$f(Y) = \text{“b”} + \text{“a”} g(X)$$

and solving by ordinary linear least squares. The AOV table which is printed out for each model is in units of the transformed Y's. R^2, the squared multiple correlation coefficient is expressed in units of the transformed Y's. The following models are provided:

Linear: $Y = aX + b$
Quadratic: $Y = aX^2 + bX + c$
Exponential: $Y = a\exp(bX)$
Power: $Y = aX \uparrow b$

Sign Test

- Object

The sign test is designed for testing whether two populations have the same medians.

- Data

The data consist of observations on a bivariate random sample (X1, Y1), ... , (Xn, Yn). Within each pair, (Xi, Yi), a comparison is made and the pair is a "+" if Xi > Yi, and a "-" if Xi < Yi. If Xi = Yi, the pairs are excluded from the analysis.

- Hypotheses

1. H0: P(Xi < Yi) = P(Xi > Yi) for all i
   H1: Either P(Xi > Yi) < P (Xi < Yi) for all i or
   P(Xi > Yi) > P(Xi < Yi) for all i
2. H0: P(Xi > Yi) ≤ P(Xi < Yi) for all i
   H1: P(Xi > Yi) > P(Xi < Yi) for all i
3. H0 = P(Xi > Yi) ≥ P(Xi < Yi) for all i
   H1 = P(Xi > Yi) < P(Xi < Yi) for all i

- Test Statistic

T = total number of pluses (+).

- Decision Rule

In this program a standardized T value Zt is printed so you can compare it to the cumulative distribution for a standardized normal random variable, Z.

1. Reject H0 if $1 - P[-Zt < Z < Zt] < \alpha$
   Accept H0 if $1 - P[-Zt < Z < Zt] > \alpha$
2. Reject H0 if $1 - P[Z \leq Zt] < 1 - \alpha$
   Accept H0 if $1 - P[Z \leq Zt] > 1 - \alpha$
3. Reject H0 if $1 - P[Z \leq Zt] > \alpha$
   Accept H0 if $1 - P[Z \leq Zt] < \alpha$

Wilcoxon Signed Ranks Test

- Object

This test is designed to test whether a particular sample came from a population with a specified median. It may also be used for paired samples to see if two samples have the same median.

- Data

The data consist of N observations (X1,Y1), (X2,Y2), ... , (XN,YN). The absolute differences $|Di| = |Xi - Yi|$, for i = 1, ..., N are computed for each pair. Ranks from 1 to N are assigned to these N pairs according to the relative size of the absolute differences. Pairs for which Xi = Yi are excluded from the analysis.

- Hypotheses

1. H0: E(X) = E(Y)
   H1: E(X) > E(Y)
2. H0: E(X) = E(Y)
   H1: E(X) < E(Y)
3. H0: E(X) = E(Y)
   H1: E(X) $\neq$ E(Y)

- Test Statistic

Define Ri = 0 if Yi > Xi (Di is negative)
Ri = the rank assigned to (Xi, Yi) if Xi > Yi

Then the test statistic $T = \Sigma Ri$, for i = 1, ..., N.

- Decision Rule

Look up the Quantiles, W(*) of the Wilcoxon signed ranks test statistic in the table included in this manual.

1. Reject H0 if $T > W(1 - \alpha)$
   Accept H0 if $T \leq W(1 - \alpha)$

2. Reject H0 if $T < W(\alpha)$
   Accept H0 if $T \geqslant W(\alpha)$
3. Reject H0 if $T > W(1-\alpha/2)$ or $T < W(\alpha/2)$
   Accept H0 if $W(\alpha/2) < T < W(1-\alpha/2)$

Higher Power Signed Rank
Ranks the N differences, Xi – Yi, from smallest to greatest. T, the test statistic, is given by the sum of the ranks of the positive differences raised to the specified power (2,3,4, or 5). Note that if the power specified were 1, this test is the Wilcoxon Signed Rank test, and if the power were 0, this test is the Sign test.

Using higher powers of the ranks can lead to a more powerful test when it is desired to weight larger values more heavily. This would be true in highly skewed distributions.

Spearman's Rho

● Object

This routine will test the independence of two random variables.

● Data

The data consist of a bivariate random sample of size N, (X1, Y1), ..., (XN, YN). Let R(Xi) be the rank of Xi as compared with the other X values, for i = 1,2, ..., N. That is R(Xi) = 1 if Xi is the smallest of X1, X2, ..., XN; R(Xi) = 2 if Xi is the second smallest, etc. Similarly, let R(Yi) equal 1,2, ..., N depending on the relative magnitude of Yi.

● Measure of Correlation

$d = \Sigma(R(X_i) - R(Y_i))^2$ for $i = 1,2,\ldots,N$
$R = 1 - [6d/N(N \uparrow 2 - 1)]$

● Hypothesis Testing

The Spearman rank correlation coefficient is used as a test statistic to test for independence between two random variables.

1. Two-Tailed Test
   H0: The Xi and Yi are mutually independent.
   H1: Either
   a) there is a tendency for the larger values of X to be paired with the larger values of Y, or
   b) there is a tendency for the smaller values of X to be paired with the larger values of Y.
2. One-Tailed Test For Positive Correlation
   H0: The Xi and Yi are mutually independent.
   H1: There is a tendency for the ranks of X and Y to be paired together.
3. One-Tailed Test For Negative Correlation
   H0: The Xi and Yi are mutually independent.
   H1: There is a tendency for the smaller values of X to be paired with the larger values of Y, and vise versa.

• Decision Rule

From the table of quantiles of the Spearman test statistic in this manual, we can find the quantile value.

1. Two-tailed test: Reject H0 if R exceeds the $(1-\alpha/2)$ quantile or if R is less than the $\alpha/2$ quantile.
2. One-tailed test for positive correlation: Reject H0 if R exceeds the $1-\alpha$ quantile.
3. One-tailed test for negative correlation: Reject H0 if R less than α quantile.

Kendall's Tau

• Object

This routine allows you to test the independence of two random variables.

• Data

The data consist of a bivariate random sample of size N, (Xi,Yi) for i = 1,2,....,N. Two observations, for example (1.3, 2.2) and (1.6,2.7), are called concordant if both members of one observation are larger than the respective members of the other observation. Pc denotes the number of concordant pairs of observations. A pair of observations like (1.3,2.2) and (1.6, 1.1) are called discordant if the two numbers in one observation differ in opposite directions (one negative and one positive) from the respective members in the other observation. Let Pd denote the number of discordant pairs of observations. If Xi = Xj or Yi = Yj, (i ≠ j), the pair is disregarded.

• Measure of Correlation

T = (Pc-Pd)/[N(N-1)/2]

• Hypotheses

Same as in Spearmans's Rho.

• Decision Rule

From the table of quantiles of the Kendall rank correlation coefficient in this manual, we can find the quantile value, Q.

1. Two-tailed test: Reject H0 if Q exceeds the $(1-\alpha/2)$ quantile or if Q is less than the $\alpha/2$ quantile.
2. One-tailed test for positive correlation: Reject H0 if Q exceeds the $1-\alpha$ quantile.
3. One-tailed test for negative correlation: Reject H0 if Q is less than the α quantile.

Two Independent Sample Tests

Object of Program

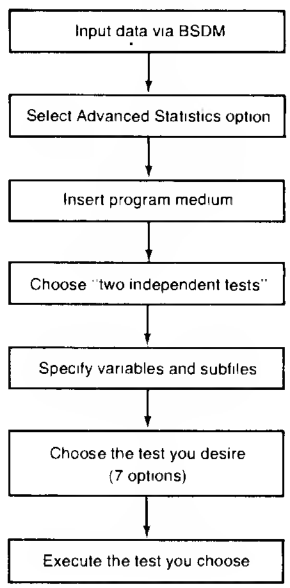
The following routines are provided:

Two-sample t-test — tests whether the means of two samples are equal.

Median test — tests whether the medians of two samples are equal.

Mann-Whitney, Taha's Squared R, Cramer-von Mises, and Kolmogorov-Smirnov tests — all test whether the two populations have the same distribution.

Typical Program Flow



Data Structure

For all of the two-independent-sample tests, data must be entered into one variable in the data base created by Basic Statistics and Data Manipulation. Then, the Subfile routine of BSDM must be used to create two subfiles. Each subfile corresponds to one sample. For example, suppose you have one sample of size six and another sample of size eight. Suppose the data is:

Sample 1: 2, 3, 4, 2, 3, 6
Sample 2: 4, 5, 4, 2, 2, 6, 3, 7.

The data should be entered vis BSDM as one variable with 14 observations. Then, the Subfile routine would be used to specify two subfiles, the first with six observations, and the second with eight observations.

Methods and Formulae

Two-Sample t Test

• Object

The two-sample t-test is used to test whether the means of two samples drawn from normal populations having the same variance are equal.

• Data

Let X1, ..., Xn be a random sample from the first population and Y1, ... , Ym be a random sample from the second. Let M(X) and M(Y) be the respective sample means and let S(X) and S(Y) be the sample variances.

• Hypotheses

Let $\mu(X)$ and $\mu(Y)$ be the two population means.

1. Two-Sided Test
   H0: $\mu(X) = \mu(Y)$
   H1: $\mu(X) \neq \mu(Y)$
2. One-Sided Test
   H0: $\mu(X) = \mu(Y)$
   H1: $\mu(X) < \mu(Y)$
3. One-Sided Test
   H0: $\mu(X) = \mu(Y)$
   H1: $\mu(X) > \mu(Y)$

• Test Statistic

$$t = [M(X) - M(Y)] / [(\frac{1}{n} + \frac{1}{m}) (\Sigma X_i \uparrow 2 - nM(X) \uparrow 2 + \Sigma Y_i \uparrow 2 - mM(Y) \uparrow 2) / [n + m - 2]^{1/2}$$

• Decision Rule

1. Two-Sided Test
   Reject H0 if $P[-t < T < t] > 1 - \alpha$
2. One-Sided Tests
   Reject H0 if $P[T<t] > 1 - \alpha$
3. One-Sided Tests
   Reject H0 if $P[T<t] < \alpha$

Median Test

• Object

The median test is designed to determine whether two samples came from populations having the same median.

• Data

From each of two populations a random sample of size Ni is obtained. Let N = N1 + N2. We obtain the sample median of the combined samples which is called the grand median. Let O1i be the number of observations in the ith sample that exceed the grand median, and let O2i be the number of observations in the ith sample that are less than or equal to the grand median. Arrange the frequency counts into a 2-by-2 contingency table as follows:

Sample	1	2	Totals
> median	O_{11}	O_{12}	
=< median	O_{21}	O_{22}	
	N_1	N_2	N

• Hypothesis

H0: The two populations have the same median.
H1: The medians of the two populations are different.

• Test Statistic

In the first sample count the number of X's greater than the grand median, say O_{11}, and the number of X's smaller than the grand median, say O_{21}, then, let $T = O_{11} - O_{21}$. The data value which is the same as the grand median is omitted.

From the contingency table, a χ^2 value can be calculated by using:

$$\chi^2 = \Sigma((O1i - O2i)^2/Ni) \text{ for } i=1,2.$$

• Decision Rule

A standardized z-value is printed, so we can look in the cumulative normal frequency distribution table to find the probability corresponding to the standardized z value, Zt, for $Z = \sqrt{\chi^2}$.

Accept H0 if $1 - P[-Zt < Z < Zt] > \alpha$
Reject H0 if $1 - P[-Zt < Z < Zt] < \alpha$

If you wish to use the χ^2 value calculated from the contingency table, then look in the chi-square contingency table and find the $W(1-\alpha)$ value with one degree of freedom where α is the significance level.

Accept H0 if calculated $\chi^2 < W(1-\alpha)$
Reject H0 if calculated $\chi^2 > W(1-\alpha)$

If $N1+N2<30$, Fisher's exact probability, P, is given. If $\alpha/2 < P < 1-\alpha/2$, accept H0; otherwise, reject H0.

Mann-Whitney Test

• Object

The Mann-Whitney test is designed to test if two populations are identical.

- Data

The data consist of two random samples. Let X1, X2, ..., XN denote the random sample of size N from population one, and let Y1, Y2, ..., YM denote the random sample of size M from population two. Assign the ranks 1 through N + M to the combined samples. Let R(Xi) and R(Yj) denote the ranks assigned to X and Y respectively, for all i and j.

- Hypotheses

Let F(X) and G(X) be the distribution functions corresponding to populations one and two respectively (or of X and Y respectively).

1. Two-Sided Test
   H0: $F(X) = G(X)$ for all X
   H1: $F(X) \neq G(X)$ for at least one X
2. One-Sided Test
   H0: $P(X < Y) \leqslant .5$
   H1: $P(X < Y) > .5$
3. One-Sided Test
   H0: $P(X < Y) \geqslant .5$
   H1: $P(X < Y) < .5$

- Test Statistic

Let $T = \Sigma\, R(Xi)$ for $i = 1, ..., N$.

In our output T is standardized to z by using:

$$z = (T - \mu)/\sigma$$

where

$$\mu = N(N + M + 1)/2$$

and

$$\sigma^2 = MN(M + N + 1)/12$$

- Decision Rule

Look in the normal probability function table to find the probability corresponding to the standardized z, Zt.

1. Two-Sided Test
   Accept H0 if $P[-Zt \leqslant Z \leqslant Zt] < 1 - \alpha$
   Reject H0 if $P[-Zt \leqslant Z \leqslant Zt] > 1 - \alpha$
2. One-Sided Test
   Accept H0 if $P[Z \leqslant Zt] > \alpha$
   Reject H0 if $P[Z \leqslant Zt] < \alpha$
3. One-Sided Test
   Accept H0 if $P[Z \leqslant Zt] < 1 - \alpha$
   Reject H0 if $P[Z \leqslant Zt] > 1 - \alpha$

Taha's Squared R

This test is similar to the Mann-Whitney test, because it ranks the pooled sample of X's and Y's and defines T by $T = \Sigma R(X_i) \uparrow 2$. Again, the null hypothesis is that the two populations have the same distribution. Z is normalized by $z = (T - \mu)/\sigma$ where

$$\mu = N(N + M + 1)(2(N + M) + 1)/6$$

and σ is very complicated, but can be found in Mielke. (See References)

Cramer-Von Mises Test

• Object

The Cramer-Von Mises test is designed to test if two populations are identical.

Data

The data consist of two independent random samples, X1, ..., XN and Y1, ..., YM, with unknown distributions functions F(*) and G(*) respectively.

• Hypothesis

H0: F(X) = G(X) for all X
H1: F(X) ≠ G(X) for at least one X

• Test Statistic

Let F1(Xi) and G1(Yj) be the empirical cumulative distribution functions. Then

$$T = \Sigma[F1(Xi) - G1(Yj)]$$

where the sum is over consecutive i and j, that is, over the "pooled" cumulative distribution function.

• Decision Rule

In the program output, T and the .10, .05, and .01 significance levels are printed. Choose your desired significance level and:

Reject H0 if T > corresponding critical point
Accept H0 is T < corresponding critical point

Kolmogorov-Smirnov Test

• Object

This test is designed to test whether two populations have the same distribution.

• Data

The data consist of two independent random samples X1, ..., XN and Y1, ..., YM. Let F(*) and G(*) represent their respective, unknown, distribution functions.

- Hypotheses
  1. Two-Sided Test
     H0: F(X) = G(X) for all X
     H1: F(X) $\neq$ G(X) for at least one value of X
  2. One-Sided Test
     H0: F(X) = G(X) for all X
     H1: F(X) > G(X) for at least one value of X
  3. One-Sided Test
     H0: F(X) = G(X) for all X
     H1: F(X) < G(X) for at least one value of X

- Test Statistic

Let S1(X) be the empirical distribution function based on the random sample X1, ..., XN, and let S2(Y) be the empirical distribution function based on the other random sample Y1, ..., YM.

Define the test statistic, T, as the greatest vertical distance between the two empirical distribution functions:

$$T = \sup | S1(X) - S2(Y) |$$

- Decision Rule

The output consists of T and the .10, .05, and .01 significance levels. Choose your desired significance level. Reject H0 if T > corresponding critical point Accept H0 otherwise

Multiple-Sample ($\geq$ 3 Samples) Tests

Description

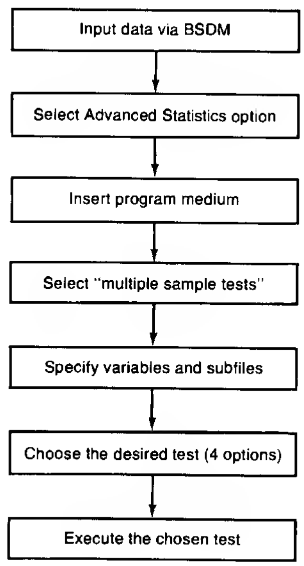
The following routines are available:

One-Way Analysis of Variance — tests whether the means of several populations are equal.

Multiple Comparisons — test whether there are significant differences between pairs of means via Least Significant Differences, Duncan's test, Student-Newman-Keul's test, Tukey's HSD, or Scheffé's test.

Kruskal-Wallis Test — tests if several populations have identical medians.

Typical Program Flow



Data Structure

For $\geq$ 3 Sample tests, three or more different subfiles of the same variable must be used. The data are entered as in the following example. Suppose you have three samples:

Sample 1: 2, 5, 8, 7, 6, 4
Sample 2: 3, 2, 9, 11
Sample 3: 7, 3, 5, 8, 6

You would enter the data via Basic Statistics and Data Manipulation as one variable with 15 observations like this:

Variable #1

I	OBS(I)	OBS(I+1)	OBS(I+2)	OBS(I+3)	OBS(I+4)
1	2	5	8	7	6
6	4	3	2	9	11
11	7	3	5	8	6

Then, the Subfile option would be used to specify three subfiles, the first with six observations, the second with four observations, and the third with five observations.

Methods and Formulae

1. **One-way Analysis of Variance** is used to test the hypothesis that the means of several populations are equal. The assumption is that all the populations are normal and have equal variances, although the sample sizes may be unequal.

   Suppose k is the number of populations and n_i is the number of observations in the sample from the ith population. The total variation of the data is

   $$SST = \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \left((X_{ij} - \bar{\bar{X}})^2 \right) \right)$$

   where $\bar{\bar{X}}$ is the overall mean. The variation due to error, or variation within samples is

   $$SSE = \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \left((X_{ij} - \bar{X}_i)^2 \right) \right)$$

   where $\bar{X}_i$ is the mean of the ith sample. The variation between samples is

   $$SSB = \sum_{i=1}^{k} (n_i (\bar{X}_i - \bar{\bar{X}})^2)$$

   The error mean square is defined as

   $$MSE = SSE/(N-k), \quad \text{where } N = \sum_{i=1}^{k} (n_i)$$

   and the between samples mean square is defined as $MSB = SSB/(k-1)$.

The F-ratio, MSB/MSE, has the F distribution with $k-1$ and $N-k$ degrees of freedom. The null hypothesis that the population means are equal may be rejected if the F ratio is greater than or equal to $F\alpha$, $k-1$, $N-k$, where α is the significance level of the experiment. This may be summarized in a table:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Between samples	$K-1$	SSB	$MSB = \frac{SSB}{k-1}$	$\frac{MSB}{MSE}$
Error	$N-k$	SSE	$MSE = \frac{SSE}{N-k}$	
Total	$N-1$	SST		

Multiple Comparisons

Multiple comparisons provide you with several tests to determine whether the the various samples have significantly different means. The procedures are used upon completion of an analysis of variance. The notation used in these tests is defined below.

EMS = error mean square used in testing for significance in the analysis of variance

n_0 = harmonic average of observations per mean
$S(M) = \sqrt{EMS/n_0}$
k = number of groups
a = degrees of freedom for EMS = n-k
Mi = mean of the ith sample, i = 1, ..., k
Oi = ith ordered (from largest to smallest) group mean, i = 1, ..., k
msd = minimum significant difference

Group means are sorted and then all possible comparisons are made. Only one table value is necessary for Least Significant Differences, Tukey's HSD, or Scheffe's test. On the other hand, k − 1 table values are needed for Student-Newman-Keul's test and Duncan's multiple range test.

The minimum significant difference is the smallest difference there can be between two means for them to be considered significantly different from one another. In all of the procedures, comparisons are made starting with the largest difference between means and progressing to the smallest difference. The process should be terminated when there is no significant difference found at a given step.

In all cases the hypothesis is:

H0: $\mu i = \mu j$, where μi is the mean of the ith population, $i \neq j$
H1: $\mu i \neq \mu j$

Least Significant Differences (Multiple Comparisons)

- Test Statistic

  msd = $t(a,b)S(M)\sqrt{2}$, where $t(a,b)$ is the upper b point of the t-distibution with a degrees of freedom

- Decision Rule

  Accept H0 if $Mi - Mj < msd$
  Reject H0 otherwise

Duncan's Multiple Range Test (Multiple Comparisons)

- Test Statistic

First, the sample means are ordered from largest to smallest: O1, O2, ..., Ok. Define p = difference in ranks of the means being compared plus one. For example, if you are comparing O2 and O5, then $p = (5 - 2) + 1 = 4$. Then:

  msd = $R(a,p,b)S(M)$, where $R(a,p,b)$ is the upper b point from the new multiple range table with a degrees of freedom and distance p.

- Decision Rule

Accept H0 if $Oi - Oj < msd$, where $i < j$
Reject H0 otherwise

Scheffe's Test (Multiple Comparisons)
After you have collected the data and tested those contrasts that catch your eye during the analysis, you should use Scheffe's Test.

- Test Statistic

  msd = $\sqrt{(k - 1)F(b,k-1,a)}\ S(M)$, where $F(b,k-1,a)$ is the upper b point of the F distributrion with $k-1$ and a degrees of freedom.

- Decision Rule

  Accept H0 if $Mi - Mj < msd$
  Reject H0 otherwise

Tukey's HSD (Multiple Comparisons)
- Test Statistic

  msd = $R(k,a,b)S(M)$, where $R(k,a,b)$ is the upper b point of the Studentized range table with a degrees of freedom and total sample number k.

- Decision Rule

  Accept H0 if $Mi - Mj < msd$
  Reject H0 otherwise

Student-Newman-Keuls Test (Multiple Comparisons)
First, the means of the sample are ordered from largest to smallest, O1, O2, ..., Ok. Then p is defined the same as in Duncan's Test.

- Test Statistic

 msd = R(p,a,b)S(M), where R(p,a,b) is the upper b point from the Studentized range table with a degrees of freedom and distance p.

- Decision Rule

 Accept H0 if msd > Oi − Oj, i < j
 Reject H0 otherwise

Kruskal-Wallis Test
- Object

The Kruskal-Wallis test is designed to test whether k independent samples, $k \geqslant 2$, have the same mean. The test does not assume normality of the k populations.

- Data

The data consist of k independent samples, each of size Ni, i = 1, ..., k. Let N = N1 + N2 + ... + Nk. Rank the combined samples. Then, for each sample compute the sum of the ranks of the observations in the sample. Call these sums Ri, for i = 1, ..., k. If more than one observation have the same value, assign the average rank to each of the tied observations.

- Hypothesis

 H0: All of the k populations have equal means
 H1: At least one of the populations has a different mean

- Test Statistic

$$T = [12/N(N+1)]\,[\Sigma(R_i \uparrow 2/N_i)] - 3(N+1), \text{ for } i = 1,\ldots,k$$

- Decision Rule

The output prints out a chi-square statistic along with the probability that a chi-square random variable is greater than the statistic. If the probability printed is smaller than the significance level you chose, reject H0. Otherwise, accept H0.

References

1. Bancroft, T.A., Topics in Intermediate Statistical Methods, Volume 1. Iowa State University Press; Ames, Iowa, 1968.
2. Boardman, T.J., and Moffitt, D.R., "Graphical Monte Carlo Type I Error Rates for Multiple Comparisons Procedures", Biometrics, 27: September 1971.
3. Conover, W.M. (1971), Practical Nonparametric Statistics. John Wiley and Sons, Inc. New York.
4. Conover, W.J. (1974), "Some Reasons For Not Using the Yates Contingency Correction on 2x2 Contingency Tables)". JASA, June 1974, 69:374.
5. Dixon, Wilfred and Massey, Frank, Introduction to Statistical Analysis, McGraw-Hill, New York, 1969, pp. 119-123.
6. Draper, N.R. and Smith, H., Applied Regression Analysis, John Wiley & Sons, New York, 1966, pp. 7-20.
7. Mielke, P.W. (1967), "Note on Some Squared Rank Tests with Existing Ties". Technometrics, 9:312.
8. Mielke, P.W. (1972), "Asymptotic Behavior of Two-Sample Tests Based on Powers of Ranks for Detecting Scales and Location Alternatives".
9. Mosteller, F. and Robert E.K. Rourke (1973), Sturdy Statistics. Addison-Wesley Publishing Co., Reading, Mass.
10. Siegel, S. (1956), Nonparametric Statistics. McGraw-Hill, New York.
11. Snedecor, George and Cochran, William, Statistical Methods, Iowa State University Press, Ames, Iowa; 1971, pp. 91-119.

Statistical Distributions

Object of Program

This program allows you to run a series of continuous and discrete statistical distributions. Both tabled values and right-tailed probabilities are available for the continuous distribution. The discrete distributions calculate right-tailed probabilities, single term probabilities and an approximate value for a specified right-tailed probability.

Additionally, this program will calculate n factorial, the complete gamma function, the complete beta function and binomial coefficients.

Methods and Formulae

Continuous

The continuous distributions included in this program are:

1. Normal (Gaussian)
2. Two-parameter gamma
3. Central F
4. Beta
5. Student's T
6. Weibull
7. Chi-square
8. Laplace (double exponential, bilateral exponential, extreme distribution, or Poisson's first law of error)
9. Logistic (autocatalytic function, growth curve)

For the central F, beta, T, chi-square and gamma distributions, the algorithms generally converge most rapidly for small or large right tail probabilities. For moderate tails, the time increases as the right tail approaches .5. For the beta distribution, both parameters should be greater than 10^{-3}. If the parameters are smaller than this, the time required for convergence is excessive.

For the chi-square, it is recommended that the degrees of freedom be less than 500.

For the logistic, Laplace and Weibull it is necessary that the right-tailed probabilities, p, satisfy $1-10^{-95}>p>10^{-95}$

For the incomplete gamma, it is recommended that the ratio A/B be less than 250.

Some special terms are:

1. **Right-tailed probability.** Given that X is a random variable and "a" is an observable value of X, then the right-tailed probability associated with "a" is PR(X>a).
2. **Tabled values.** Given that X is a random variable and P is a right-tailed probability, then the tabled value associated with P is that value "a" such that PR(X>a) = P.

To specify the distributions, the respective density functions that are evaluated will be shown below. Let f(x) be a density, and $\Gamma(*)$ be the gamma function.

1. Normal (standard)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty<x<\infty$$

2. Two parameter gamma, parameters A,B

$$f(x) = \frac{1}{\Gamma(A)B^A} * x^{A-1} * e^{-x/B} \qquad x>0 \qquad A>0,\ B>0$$

3. Central F with N degrees of freedom in the numerator and D in the denominator

$$f(x) = \frac{\Gamma((N+D)/2)(N/D)^{N/2}}{\Gamma(N/2)\,\Gamma(D/2)} \frac{x^{N/2-1}}{\left(1+\frac{Nx}{D}\right)^{(N+D)/2}} \qquad \text{N and D are positive integers}$$

4. Beta with parameters A and B

$$f(x) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}(1-x)^{B-1}x^{A-1} \qquad 0\leq x\leq 1 \qquad A,B>0$$

5. Student's t with N degrees of freedom

$$f(x) = \frac{\Gamma((N+1)/2)}{\sqrt{N\pi}\,\Gamma(N/2)} * \frac{1}{(1+x^2/N)^{(N+1)/2}} \qquad -\infty<x<\infty \qquad \text{N positive integer}$$

6. Weibull with parameters A,B

$$f(x) = BA^Bx^{B-1}\exp[-Ax^B] \qquad x>0 \qquad A,B>0$$

7. Chi-square with N degrees of freedom

$$f(x) = \frac{1}{\Gamma(N/2)\,2^{N/2}} \quad x^{N/2-1}\,e^{-x/2}$$

N is a positive integer
X>0

8. Logistic with parameters A,B

$$f(x) = \frac{Bx\exp(-(A+Bx))}{[1+\exp(-(A+Bx))]^2} \qquad B>0 \text{ and } -\infty<x<\infty$$

9. Laplace with parameters A and B

$$f(x) = \frac{1}{2B} \exp\{-|x-A|/B\} \quad B>0 \text{ and } -\infty<x<\infty$$

Discrete

The discrete distributions included in this program are:

1. Binomial
2. Negative Binomial
3. Poisson
4. Hypergeometric
5. Gamma Function
6. Beta Function
7. Single Term Binomial
8. Single Term Negative Binomial
9. Single Term Poisson
10. Single Term Hypergeometric

Other routines of this program are N factorial and Binomial Coefficients.

Some special terms used are:

1. **Tabled value.** Let X be a binomial, bypergeometric or Poisson random variable. Given all approriate parameters and p, a desired right-tailed probability, then the tabled value is defined to be x such that $P(X>x)=p$.
2. **Single term probability.** Given that X is one of the three distributions and x is the counter domain of X, then the single term probability is defined to be $P(X=x)$.

All tabled values are normal approximations. It should be noted that if a right-tailed probability p is desired, it is an unlikely coincidence that there will exist an element x in the counter domain such that $P(X>x)=p$ where x is one of the distributions in (2) above. Thus, after getting the normal approximation to the tabled value, values in the counter domain near the approximation should be checked to see which value is best for the particular application.

The distributions are defined as follows:

1. Hypergeometric

Let N = number of items in a lot	$M \le N$
M = sample size	$K \le N$
X = number of defective items in the sample	$X \le K$
K = number of defective items in the lot	$X \le M$

then P (exactly x defectives are in the sample) is

$$P(X=x)=\frac{\binom{K}{x}\binom{N-K}{M-x}}{\binom{N}{M}} \qquad ,x=0,1,\ldots,M$$

and

$$P=P(X\geqslant x)=\sum_{i=x}^{\min(M,K)} P(X=i)$$

2. Binomial

Let N = number of trials
p = probability of success at each trial
X = number of successes

$$P(X=R)=\binom{N}{R}p^R(1-p)^{N-R}, \qquad R=0,1,\ldots,N\,,\ 0<p<1$$

and

$$P=\ P(X\geqslant R)=\sum_{i=R}^{N}\binom{N}{i}p^i(1-p)^{N-i}$$

3. Poisson

Let m = rate parameter or mean = lambda >0
X = number of occurrences = 0,1,2,...

$$P=P(X\geqslant N)=e^{-m}\sum_{i=N}^{\infty}\frac{m^i}{i!}$$

4. Negative Binomial

For a sequence of Bernoulli trials with probability p of success,

let R = number of failures before the Nth success then

$$P(X=R)=\binom{N+R-1}{R}p^N(1-p)^R, \qquad R=0,1,2\ldots,\ 0<p<1$$

and if A = number of failures before the Nth success then

$$P(X \geq A) = \sum_{i=A}^{\infty} \binom{N+i-1}{i} p^N (1-p)^i, \qquad A=0,1,2$$

5. N! and $\Gamma(x)$ and complete beta function. N must be a non-negative integer.

   An asymptotic Stirling's approximation is used to calculate N! and $\Gamma(x)$ and complete beta function.

Special Considerations

Loading the Program Directly

This program may be entered via Basic Statistics and Data Manipulation, any One Sample test, or any Multiple Sample test. You may also load the program directly by following these instructions:

1. Insert the General Statistics program medium.
2. Enter: LOAD "START_DIST",10,
3. Press: EXECUTE

Before you load the program directly, you must specify the mass storage device which contains the program medium using the MASS STORAGE IS command.

Continuity Correction

For right-tailed probabilities, the exact probabilities are calculated. Thus, there is no need to use a continuity correction. There is no restriction that the parameters be integers, so if for some reason a continuity correction is desired, one may be used.

References

1. Abramowitz, M. and Stegun, I.A., Handbook of Mathematical Functions, National Bureau of Standards, 1964.
2. Abramowitz, M. and Stegun, I. (1964) N.B.S. Handbook Series 55, Government Printing Office.
3. Erdelyi, A., editor (1953) Higher Transcendental Functions, Vo. 1, McGraw-Hill, New York.
4. Johnson, N., and Kotz, S. (1970) Continuous Univariate Distributions, Vol. 1 and 2, Houghton-Mifflin, New York.
5. Khovanskii, A.N., (1956) The Applications of Continued Fractions and Their Generation to Problems in Approximation Theory, P. Noordhoff, Groningen.
6. Kopitzke, R., PH.D. Dissertation, 1974.
7. Kopitzke, Robert W., Unpublished research notes.
8. Lieberman, G.J. and Own, D.B., Tables of the Hypergeometric Probability Distribution, Stanford University Press, 1961.
9. Wall, H.S., (1948) Analytic Theory of Continued Fractions, D. Van Nostrand, New York.
10. Whitaker, E.T., and Watson, G.N., (1940) Modern Analysis, Cambridge University Press.

Examples

Examples On One Sample Data Sets

One Hundred Failure-Time Data

One hundred observations of the time until failure of an electronic circuit were obtained from a life testing experiment. The coded data values are shown below. The serial correlations with lag 1 and lag 2 were quite small indicating apparent "independence" of the observations. Also, a serial plot of the data shows no particular patterns. The runs test further confirms the randomness of the data.

This type of data is assumed to come from an exponential random variable with mean = 1. The histogram of the data indicates that this assumption might be valid. If the data really is exponential with mean = 1, then the sample mean and standard deviation also should be about 1. From the output we see that x = 1.0856 and s = .9301 which do not differ from 1 by a great deal. This is confirmed by the one-sample t-test.

Both the Chi-square goodness of fit test and the Kolmogorov-Smirnov goodness of fit test indicate that we cannot reject the hypothesis that the data came from an exponentially distributed population with mean = 1. The χ^2 test yields a test statistic of 9.248 with 8 degrees of freedom, which is not significant even at the $\alpha = .10$ level. The K-S test statistic DN = .09907, is not significant at $\alpha = .20$ level. However, both tests (χ^2 and K-S) indicate that the data is not normally distributed.

Since the sample size for this example was too large to perform a Shapiro Wilk Normality test, half of the observations were selected to give you an idea of the output.

```
*****************************************************************************************
*                                    DATA MANIPULATION                                  *
*****************************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                                  Raw data
Mode number = ?
2                                                                  On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
TIME:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES




Data file name: TIME:INTERNAL

Data type is:   Raw data

Number of observations:    100
Number of variables:         1
```

Variable names:
1. X1

Subfiles: NONE

SELECT ANY KEY

Option number = ?
1

Press special function key labeled-LIST

Data type is: Raw data

VARIABLE # 1 (X1)

I	OBS(I)	OBS(I+1)	OBS(I+2)	OBS(I+3)	OBS(I+4)
1	2.00790	2.45450	2.55760	.50250	1.71430
6	1.71430	2.52480	.84390	2.89900	.32220
11	.18180	3.38780	1.71490	.16020	.10360
16	.53530	1.18870	.01480	.03510	.21580
21	.84770	1.85770	1.08500	3.25370	1.73570
26	1.03880	1.72300	1.72300	1.85580	.89840
31	.14220	.12790	1.49950	.11010	3.37350
36	.60190	1.90800	.52140	.29580	.49730
41	1.63010	.05740	1.08360	.57650	2.25210
46	2.72780	.83400	1.14640	.02070	.23900
51	3.84480	1.29530	.81290	.85020	.97390
56	.43280	.83970	1.08490	.95980	.51170
61	.89530	2.51070	.32380	1.06270	3.21960
66	1.20550	.39400	.29730	1.27110	.98670
71	2.31500	.48060	1.34410	.78670	2.28790
76	.12190	.54020	3.11250	.17480	.06320
81	.65310	.54450	.01050	.18050	.46430
86	.55340	.99490	.28950	1.36600	.15090
91	1.51270	1.53900	.77450	.14300	.44900
96	.43340	.16540	1.76060	.40100	.43230

Option number = ?
0

Exit LIST procedure

SELECT ANY KEY

Select special function key labeled ADV. STAT
Remove BSDM media
Insert General Statistics media

Enter number of desired function:
1

Choose 1 sample tests

****

ONE SAMPLE TESTS

VARIABLE --X1

****

Enter desired function:
1

Choose serial correlation

SERIAL CORRELATION SAMPLE SIZE IS 100

CORRELATION LAG = ?
1

Choose lag = 1

SERIAL CORRELATION WITH LAG = 1 IS .01605

Not very serially correlated

```
ENTER ANOTHER LAG?
YES
CORRELATION LAG = ?
2
     SERIAL CORRELATION WITH LAG =  2 IS  -.01235
```

Try lag = 2
Not very correlated

```
ENTER ANOTHER LAG?
NO

Enter desired function:
2
```

Obtain ranks

```
RANKED DATA:
------------
               DISTINCT                          DISTINCT                          DISTINCT
(    RANK     DATA POINT)  (    RANK     DATA POINT)  (    RANK     DATA POINT)

(     1.00        .0105)  (     2.00        .0148)  (     3.00        .0207)
(     4.00        .0351)  (     5.00        .0574)  (     6.00        .0632)
(     7.00        .1036)  (     8.00        .1101)  (     9.00        .1219)
(    10.00        .1279)  (    11.00        .1422)  (    12.00        .1430)
(    13.00        .1509)  (    14.00        .1602)  (    15.00        .1654)
(    16.00        .1748)  (    17.00        .1805)  (    18.00        .1818)
(    19.00        .2158)  (    20.00        .2390)  (    21.00        .2895)
(    22.00        .2958)  (    23.00        .2973)  (    24.00        .3222)
(    25.00        .3238)  (    26.00        .3940)  (    27.00        .4010)
(    28.00        .4323)  (    29.00        .4328)  (    30.00        .4334)
(    31.00        .4490)  (    32.00        .4643)  (    33.00        .4806)
(    34.00        .4973)  (    35.00        .5025)  (    36.00        .5117)
(    37.00        .5214)  (    38.00        .5353)  (    39.00        .5402)
(    40.00        .5445)  (    41.00        .5534)  (    42.00        .5765)
(    43.00        .6019)  (    44.00        .6531)  (    45.00        .7745)
(    46.00        .7867)  (    47.00        .8129)  (    48.00        .8340)
(    49.00        .8397)  (    50.00        .8439)  (    51.00        .8477)
(    52.00        .8502)  (    53.00        .8953)  (    54.00        .8984)
(    55.00        .9598)  (    56.00        .9739)  (    57.00        .9867)
(    58.00        .9949)  (    59.00       1.0388)  (    60.00       1.0627)
(    61.00       1.0836)  (    62.00       1.0849)  (    63.00       1.0850)
(    64.00       1.1464)  (    65.00       1.1887)  (    66.00       1.2055)
(    67.00       1.2711)  (    68.00       1.2953)  (    69.00       1.3441)
(    70.00       1.3660)  (    71.00       1.4995)  (    72.00       1.5127)
(    73.00       1.5390)  (    74.00       1.6301)  (    75.50       1.7143)
(    77.00       1.7149)  (    78.50       1.7230)  (    80.00       1.7357)
(    81.00       1.7606)  (    82.00       1.8558)  (    83.00       1.8577)
(    84.00       1.9080)  (    85.00       2.0079)  (    86.00       2.2521)
(    87.00       2.2879)  (    88.00       2.3150)  (    89.00       2.4545)
(    90.00       2.5107)  (    91.00       2.5248)  (    92.00       2.5576)
(    93.00       2.7278)  (    94.00       2.8990)  (    95.00       3.1125)
(    96.00       3.2196)  (    97.00       3.2537)  (    98.00       3.3735)
(    99.00       3.3878)  (   100.00       3.8448)


Enter desired function:
3
```

Choose t-test

```
ONE-SAMPLE t-TEST    SAMPLE SIZE IS  100
-----------------

1 OR 2 TAIL TEST
2
```

2 tail test

```
                2 TAIL TEST
H0:  MU= 1.085611  OR =
?
```

```
1.0000                                                     Specify hypothesis mean
              H0:  MU=       1

     N=                         100
     MEAN=                        1.0856
     STD DEV =                     .9301
     STD ERROR OF MEAN=            .0930
     t=                            .9204
     DF=                         99                        Cannot reject hypothesis


     1 - P(    -.9204  < t <     .9204) =         .3596


Enter desired function:
4                                                          Choose Kolmogorov-Smirnov G.O.F. test

KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST    SAMPLE SIZE IS  100
---------------------------------------

Please enter G.O.F. code:
2                                                          Choose exponential form of the
                                                           hypothesized distribution.

Testing for  EXPONENTIAL  goodness of fit.


MEAN= 1.085611 OR=
?
1
    MEAN = 1

    N= 100,  KOLMOGOROV-SMIRNOV STATISTICS:  DN =          .09907
                                       SQR(N)*DN =          .99


ANOTHER G.O.F. CODE?
NO

Enter desired function:
5                                                          Choose Chi-square G.O.F. test

CHI-SQUARE GOODNESS-OF-FIT TEST    SAMPLE SIZE IS  100
-------------------------------
Please enter G.O.F. code:
2                                                          Select exponential distribution again

Testing for  EXPONENTIAL  goodness of fit.


OFFSET =
0                                                          Minimum value for histogram
    OFFSET = 0
# OF CELLS (max is 50) = ?
10                                                         10 intervals or windows
    # OF CELLS = 10
    OPTIMUM CELL WIDTH =    .3845

CELL WIDTH = .3844838448   OR =
?
.4
```

YOUR CELL WIDTH = .4000

CELL #	LOWER LIMIT	OBSERVED # OF OBS.	EXPECTED # OF OBS.
1	0.0000	26	30.82
2	.4000	20	21.32
3	.8000	19	14.75
4	1.2000	8	10.20
5	1.6000	11	7.06
6	2.0000	4	4.88
7	2.4000	5	3.38
8	2.8000	2	2.34
9	3.2000	4	1.62
10	3.6000	1	1.12

```
    CHI-SQUARE GOODNESS-OF-FIT FOR EXPONENTIAL DISTRIBUTION
    CHI-SQUARE VALUE =    9.248;   DEGREES OF FREEDOM =   8
```
Not very big.

```
ANOTHER GOF CODE?
NO
```
See Chi-square table in appendix with 8 degrees of freedom.

```
Enter desired function:
7
```
Choose runs test

```
RUNS TEST    SAMPLE SIZE IS  100
---------

Select a significance level by entering 1, 2 or 3:
3
```
Choose α = .05

```
TEST FOR TOO FEW RUNS?
YES
```
See if data is too non-random

```
    # OF RUNS IS NOT SIGNIFICANT AT THE  .05
    SIGNIFICANCE LEVEL FOR TOO FEW RUNS
TEST FOR TOO MANY RUNS?
NO
Another significance level?
NO

Enter desired function:
9
```
Exit one-sample tests

```
Enter number of desired function:
6
```
Return to BSDM to split data set in half for Shapiro-Wilk test.

```
SELECT ANY KEY
```
Select special function key labeled-SUBFILES

```
Option number = ?
1
Number of subfiles ( <=20 ) = ?
2
```
Split data set by specifying number of observations in each subfile

```
Name of Subfile # 1 ( <=10 characters ) =
?
FIRST HALF
Subfile # 1 ; number of observations =
?
50
Name of Subfile # 2 ( <=10 characters ) =
?
SECONDHALF
Is the above information correct?
YES
Subfile name:    beginning observation  number of observations
 1 FIRST HALF                1                      50
 2 SECONDHALF               51                      50
```

```
Option number = ?
0                                                      Exit subfiles procedure
PROGRAM NOW UPDATING SCRATCH DATA FILE
SELECT ANY KEY
                                                       Return to General Statistics by pressing
                                                       ADV. STAT key
Enter number of desired function:
1                                                      Choose one-sample tests

SUBFILE NUMBER? (0=IGNORE SUBFILES)
1
*******************************************************************************

                              ONE SAMPLE TESTS

VARIABLE --X1
    SUBFILE --FIRST HALF

*******************************************************************************

Enter desired function:
6                                                      Select Shapiro-Wilk test for subfile 1


SHAPIRO-WILK NORMALITY TEST    SAMPLE SIZE IS   50
---------------------------

    W STATISTIC FOR NORMALITY =       .904821834706

                  % POINTS FOR W (SMALL VALUE SIGNIFICANT)

                             .01       .02       .05       .1       .5
    CORRESPONDING W VALUES:  .93       .938      .947      .955     .974

Enter desired function:
8

SUBFILE NUMBER? (0=IGNORE SUBFILES)
2
*******************************************************************************

                              ONE SAMPLE TESTS

VARIABLE --X1
    SUBFILE --SECONDHALF

*******************************************************************************

Enter desired function:
6                                                      Select Shapiro-Wilk test for subfile 2


SHAPIRO-WILK NORMALITY TEST    SAMPLE SIZE IS   50
---------------------------

    W STATISTIC FOR NORMALITY =       .831574211967

                  % POINTS FOR W (SMALL VALUE SIGNIFICANT)

                             .01       .02       .05       .1       .5
    CORRESPONDING W VALUES:  .93       .938      .947      .955     .974

Enter desired function:
9                                                      Return to main menu

Enter number of desired function:
6                                                      Return to BSDM
```

```
SELECT ANY KEY
```

Examples On Two Paired Samples Data Sets

Pig Weight Changes

176 pigs were paired on the basis of sex, age, and initial weight. They were fed daily one of two iron compounds to supplement that which they lacked due to confinement in pens. It was desired to determine if there was any difference in pig weight due to the two different compounds as applied over a one month period. From the paired-t test and the correlation coefficient, we see the difference is not significant.

```
*******************************************************************************
*                              DATA MANIPULATION                              *
*******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                              Raw data
Mode number = ?
2                                                              On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
PIGS:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


                         PIG WEIGHT CHANGES


Data file name: PIGS:INTERNAL

Data type is:   Raw data

Number of observations:    88
Number of variables:        2


Variable names:
   1. VARIABLE#1                                               Clever names for variables
   2. VARIABLE#2

Subfiles:  NONE


SELECT ANY KEY

Option number = ?
1                                                              List all the data
Enter method for listing data:
3
```

PIG WEIGHT CHANGES

Data type is: Raw data

OBS#	Variable # 1 (VARIABLE#1)	Variable # 2 (VARIABLE#2)
1	54.00000	46.00000
2	44.00000	42.00000
3	46.00000	44.00000
4	54.00000	44.00000
5	45.00000	45.00000
6	46.00000	52.00000
7	50.00000	51.00000
8	43.00000	55.00000
9	47.00000	60.00000
10	40.00000	43.00000
11	40.00000	20.00000
12	46.00000	48.00000
13	52.00000	54.00000
14	50.00000	55.00000
15	54.00000	62.00000
16	49.00000	41.00000
17	30.00000	48.00000
18	50.00000	45.00000
19	48.00000	46.00000
20	38.00000	31.00000
21	27.00000	35.00000
22	50.00000	59.00000
23	107.00000	135.00000
24	77.00000	90.00000
25	91.00000	98.00000
26	88.00000	98.00000
27	93.00000	96.00000
28	89.00000	74.00000
29	95.00000	98.00000
30	105.00000	133.00000
31	107.00000	126.00000
32	95.00000	91.00000
33	114.00000	52.00000
34	128.00000	98.00000
35	110.00000	119.00000
36	104.00000	105.00000
37	94.00000	110.00000
38	87.00000	81.00000
39	66.00000	83.00000
40	96.00000	112.00000
41	120.00000	104.00000
42	90.00000	101.00000
43	95.00000	88.00000
44	86.00000	86.00000
45	158.00000	221.00000
46	125.00000	176.00000
47	149.00000	150.00000
48	175.00000	176.00000
49	196.00000	209.00000
50	121.00000	118.00000
51	181.00000	180.00000
52	201.00000	238.00000
53	175.00000	196.00000
54	147.00000	138.00000
55	209.00000	133.00000
56	194.00000	159.00000
57	203.00000	209.00000
58	179.00000	205.00000

59	170.00000	201.00000
60	148.00000	149.00000
61	138.00000	159.00000
62	232.00000	230.00000
63	223.00000	198.00000
64	151.00000	161.00000
65	142.00000	147.00000
66	167.00000	176.00000
67	210.00000	320.00000
68	240.00000	267.00000
69	245.00000	221.00000
70	263.00000	247.00000
71	263.00000	293.00000
72	182.00000	211.00000
73	261.00000	178.00000
74	280.00000	320.00000
75	264.00000	266.00000
76	187.00000	178.00000
77	280.00000	199.00000
78	287.00000	230.00000
79	230.00000	256.00000
80	234.00000	272.00000
81	238.00000	245.00000
82	202.00000	222.00000
83	202.00000	245.00000
84	317.00000	243.00000
85	293.00000	264.00000
86	215.00000	215.00000
87	171.00000	172.00000
88	242.00000	233.00000

```
Option number = ?
0                                                  Exit list procedure
SELECT ANY KEY                                     Select special function key labeled-ADV. STAT
                                                   Remove BSDM media
Enter number of desired function:                  Insert General Statistics
3                                                  Choose two paired sample analyses

VARIABLE NUMBER FOR X =?
1
VARIABLE NUMBER FOR Y =?
2
*******************************************************************************
                         PAIRED SAMPLE TESTS

VARIABLE FOR X --   VARIABLE#1
VARIABLE FOR Y --   VARIABLE#2

*******************************************************************************

Enter desired function:
1                                                  Choose paired t-test

PAIRED-t TEST     SAMPLE SIZE IS    88
-------------
1 OR 2 TAILED?
1
H0 : MU(X)-MU(Y) =
0                                                  Specify zero difference
```

```
  1     TAILED TEST
 H0 : MU(X)-MU(Y) = 0
 H1 : MU(X)-MU(Y) < 0

LEVEL OF SIGNIFICANCE
.05
 T VALUE =      -.736
 DF =    87

 T(0.9500,   87 ) =      1.663

           DO NOT REJECT H0 AT  .05 LEVEL OF SIGNFICANCE

ANOTHER PAIRED-t TEST ON THIS DATA?
NO

Enter desired function:
2


CROSS CORRELATION    SAMPLE SIZE IS    88
-----------------

LAG ON X OR Y?
Y
LAG ON Y=
?
1
  LAG ON  Y =  1    COEFF. =  .85126

  ANOTHER CROSS CORRELATION?
YES
LAG ON X OR Y?
Y
LAG ON Y=
?
2
  LAG ON  Y =  2    COEFF. =  .82534

  ANOTHER CROSS CORRELATION?
YES
LAG ON X OR Y?
Y
LAG ON Y=
?
3
  LAG ON  Y =  3    COEFF. =  .88230

  ANOTHER CROSS CORRELATION?
YES
LAG ON X OR Y?
Y
LAG ON Y=
?
22
  LAG ON  Y = 22    COEFF. =  .89051

  ANOTHER CROSS CORRELATION?
NO

Enter desired function:
3


FAMILY REGRESSION / AOV    SAMPLE SIZE IS    88
-----------------------
```

Specify α = .05

Choose cross correlation

Lag of 1 on y

Try lag of 2

Try lag of 3

Try lag of 22

Choose family regression

```
REGRESSION CODE =?
1
```

Choose linear regression
Y = A + BX + E

AOV OF LINEAR REGRESSION
Y = A + BX

SOURCE	SS	DF	MS	F RATIO
REG	481475.711	1	481475.711	581.18
RES	71246.789	86	828.451	
TOTAL COR	552722.500	87		

R SQUARED = .8711

YHAT = (10.129409002) + (.943467866544)X

```
EVALUATE Y AT X ?
YES
AT ALL X(I)'S ?
YES
```

Table of predicted values and residuals

Y EVALUATED AT X

	X(I)	YHAT	Y(I)	RES(I)
1	54.000	61.0767	46.00000	15.07667
2	44.000	51.6420	42.00000	9.64200
3	46.000	53.5289	44.00000	9.52893
4	54.000	61.0767	44.00000	17.07667
5	45.000	52.5855	45.00000	7.58546
6	46.000	53.5289	52.00000	1.52893
7	50.000	57.3028	51.00000	6.30280
8	43.000	50.6985	55.00000	4.30147
9	47.000	54.4724	60.00000	5.52760
10	40.000	47.8681	43.00000	4.86812
11	40.000	47.8681	20.00000	27.86812
12	46.000	53.5289	48.00000	5.52893
13	52.000	59.1897	54.00000	5.18974
14	50.000	57.3028	55.00000	2.30280
15	54.000	61.0767	62.00000	.92333
16	49.000	56.3593	41.00000	15.35933
17	30.000	38.4334	48.00000	9.56656
18	50.000	57.3028	45.00000	12.30280
19	48.000	55.4159	46.00000	9.41587
20	38.000	45.9812	31.00000	14.98119
21	27.000	35.6030	35.00000	.60304
22	50.000	57.3028	59.00000	1.69720
23	107.000	111.0805	135.00000	23.91953
24	77.000	82.7764	90.00000	7.22357
25	91.000	95.9850	98.00000	2.01502
26	88.000	93.1546	98.00000	4.84542
27	93.000	97.8719	96.00000	1.87192
28	89.000	94.0980	74.00000	20.09805
29	95.000	99.7589	98.00000	1.75886
30	105.000	109.1935	133.00000	23.80647
31	107.000	111.0805	126.00000	14.91953
32	95.000	99.7589	91.00000	8.75886
33	114.000	117.6847	52.00000	65.68475
34	128.000	130.8933	98.00000	32.89330
35	110.000	113.9109	119.00000	5.08913
36	104.000	108.2501	105.00000	3.25007
37	94.000	98.8154	110.00000	11.18461
38	87.000	92.2111	81.00000	11.21111
39	66.000	72.3983	83.00000	10.60171
40	96.000	100.7023	112.00000	11.29768

41	120.000	123.3456	104.00000	19.34555
42	90.000	95.0415	101.00000	5.95848
43	95.000	99.7589	88.00000	11.75886
44	86.000	91.2676	86.00000	5.26765
45	158.000	159.1973	221.00000	61.80267
46	125.000	128.0629	176.00000	47.93711
47	149.000	150.7061	150.00000	.70612
48	175.000	175.2363	176.00000	.76371
49	196.000	195.0491	209.00000	13.95089
50	121.000	124.2890	118.00000	6.28902
51	181.000	180.8971	180.00000	.89709
52	201.000	199.7665	238.00000	38.23355
53	175.000	175.2363	196.00000	20.76371
54	147.000	148.8192	138.00000	10.81919
55	209.000	207.3142	133.00000	74.31419
56	194.000	193.1622	159.00000	34.16218
57	203.000	201.6534	209.00000	7.34661
58	179.000	179.0102	205.00000	25.98984
59	170.000	170.5189	201.00000	30.48105
60	148.000	149.7627	149.00000	.76265
61	138.000	140.3280	159.00000	18.67203
62	232.000	229.0140	230.00000	.98605
63	223.000	220.5227	198.00000	22.52274
64	151.000	152.5931	161.00000	8.40694
65	142.000	144.1018	147.00000	2.89815
66	167.000	167.6885	176.00000	8.31146
67	210.000	208.2577	320.00000	111.74234
68	240.000	236.5617	267.00000	30.43830
69	245.000	241.2790	221.00000	20.27904
70	263.000	258.2615	247.00000	11.26146
71	263.000	258.2615	293.00000	34.73854
72	182.000	181.8406	211.00000	29.15944
73	261.000	256.3745	178.00000	78.37452
74	280.000	274.3004	320.00000	45.69959
75	264.000	259.2049	266.00000	6.79507
76	187.000	186.5579	178.00000	8.55790
77	280.000	274.3004	199.00000	75.30041
78	287.000	280.9047	230.00000	50.90469
79	230.000	227.1270	256.00000	28.87298
80	234.000	230.9009	272.00000	41.09911
81	238.000	234.6748	245.00000	10.32524
82	202.000	200.7099	222.00000	21.29008
83	202.000	200.7099	245.00000	44.29008
84	317.000	309.2087	243.00000	66.20872
85	293.000	286.5655	264.00000	22.56549
86	215.000	212.9750	215.00000	2.02500
87	171.000	171.4624	172.00000	.53759
88	242.000	238.4486	233.00000	5.44863

```
 REGRESSION CODE =?
0
```

Exit family regression

```
Enter desired function:
10
```

Exit two-paired sample test.

```
Enter number of desired function:
6
```

Return to BSDM

Bus Passenger Service Time

The time required to service passengers boarding at a bus stop was measured together with the actual number of passengers boarding. The service time as recorded from the moment that the bus stopped and the door opened until the last passenger boarded t[illegible]us. The objective is to determine a model for predicting passenger service time, given [illegible]dge of the number boarding at a particular stop. Let X = number boarding and Y [illegible]enger service time. The following data was gathered during the month of May, 1[illegible] twelve downtown locations in Louisville, Kentucky.

```
******************************************************************************
*                              DATA MANIPULATION                             *
******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                   Raw data
Mode number = ?
2                                                   From mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
BUSTIME:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES




     BUS PASSENGER SERVICE TIME


Data file name: BUSTIME:INTERNAL

Data type is:   Raw data

Number of observations:    31
Number of variables:        2


Variable names:
   1. NUMBER
   2. TIME

Subfiles:  NONE


SELECT ANY KEY
                                                    Choose special function key labeled-LIST
Option number = ?
1                                                   List all data
Enter method for listing data:
3
```

```
 BUS PASSENGER SERVICE TIME

Data type is: Raw data


       Variable # 1   Variable # 2
       (NUMBER    )   (TIME      )

OBS#
   1        1.00000        1.40000
   2        1.00000        2.80000
   3        1.00000        3.00000
   4        1.00000        1.80000
   5        1.00000        2.00000
   6        2.00000        4.70000
   7        2.00000        8.00000
   8        2.00000        3.00000
   9        2.00000        2.50000
  10        3.00000        5.20000
  11        3.00000        6.20000
  12        3.00000        9.40000
  13        4.00000       11.70000
  14        5.00000        7.50000
  15        5.00000       11.90000
  16        6.00000       13.60000
  17        6.00000       12.40000
  18        6.00000       11.60000
  19        7.00000       14.70000
  20        7.00000       13.50000
  21        8.00000       12.00000
  22        8.00000       14.10000
  23        8.00000       26.00000
  24        9.00000       19.00000
  25       10.00000       21.20000
  26       11.00000       22.90000
  27       11.00000       22.60000
  28       13.00000       25.20000
  29       17.00000       33.50000
  30       19.00000       33.70000
  31       25.00000       54.20000


Option number = ?
0                                                   Exit list procedure
SELECT ANY KEY                                      Choose special function key labeled-ADV. STAT
                                                    Remove BSDM media
                                                    Insert General Statistics media
Enter number of desired function:
3                                                   Choose two paired sample test

VARIABLE NUMBER FOR X =?
1
VARIABLE NUMBER FOR Y =?
2
*******************************************************************************
                              PAIRED SAMPLE TESTS

VARIABLE FOR X --   NUMBER
VARIABLE FOR Y --   TIME

*******************************************************************************


Enter desired function:
3                                                   Choose family regression
```

```
FAMILY REGRESSION / AOV    SAMPLE SIZE IS    31
--------------------------

 REGRESSION CODE =?
1
```

Linear regression
Y = A + BX + E

```
                         AOV OF LINEAR REGRESSION
                               Y = A + BX

    SOURCE              SS          DF           MS          F RATIO

     REG            3970.237         1        3970.237        543.72
     RES             211.758        29           7.302
     TOTAL COR      4181.995        30

        R SQUARED =            .9494
```

Not bad!

```
        YHAT = ( .586330097087 ) + ( 1.99576699029 )X

EVALUATE Y AT X ?
YES
AT ALL X(I)'S ?
YES
                          Y EVALUATED AT X

              X(I)          YHAT          Y(I)          RES(I)

     1        1.000        2.5821        1.40000        1.18210
     2        1.000        2.5821        2.80000         .21790
     3        1.000        2.5821        3.00000         .41790
     4        1.000        2.5821        1.80000         .78210
     5        1.000        2.5821        2.00000         .58210
     6        2.000        4.5779        4.70000         .12214
     7        2.000        4.5779        8.00000        3.42214
     8        2.000        4.5779        3.00000        1.57786
     9        2.000        4.5779        2.50000        2.07786
    10        3.000        6.5736        5.20000        1.37363
    11        3.000        6.5736        6.20000         .37363
    12        3.000        6.5736        9.40000        2.82637
    13        4.000        8.5694       11.70000        3.13060
    14        5.000       10.5652        7.50000        3.06517
    15        5.000       10.5652       11.90000        1.33483
    16        6.000       12.5609       13.60000        1.03907
    17        6.000       12.5609       12.40000         .16093
    18        6.000       12.5609       11.60000         .96093
    19        7.000       14.5567       14.70000         .14330
    20        7.000       14.5567       13.50000        1.05670
    21        8.000       16.5525       12.00000        4.55247
    22        8.000       16.5525       14.10000        2.45247
    23        8.000       16.5525       26.00000        9.44753
    24        9.000       18.5482       19.00000         .45177
    25       10.000       20.5440       21.20000         .65600
    26       11.000       22.5398       22.90000         .36023
    27       11.000       22.5398       22.60000         .06023
    28       13.000       26.5313       25.20000        1.33130
    29       17.000       34.5144       33.50000        1.01437
    30       19.000       38.5059       33.70000        4.80590
    31       25.000       50.4805       54.20000        3.71950

 REGRESSION CODE =?
0
```

Exit family regression

```
Enter desired function:
10                                                    Exit two paired sample tests

Enter number of desired function:
6                                                     Return to BSDM
```

Example #3

This example is included for your convenience as a sample problem so that you may check your operation of the routines involved.

```
*******************************************************************************
*                            DATA MANIPULATION                                *
*******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                     Raw data
Mode number = ?
2                                                     On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
TWONP:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


                 TWO SAMPLE NONPARAMETRIC STATISTICS


Data file name: TWONP:INTERNAL

Data type is:   Raw data

Number of observations:    12
Number of variables:        2


Variable names:
   1. X(I)
   2. Y(I)

Subfiles:  NONE


SELECT ANY KEY
                                                      Select special function key labeled-LIST
Option number = ?
1                                                     List all data
Enter method for listing data:
3
```

```
                     TWO SAMPLE NONPARAMETRIC STATISTICS

Data type is: Raw data

          Variable # 1    Variable # 2
          (X(I)      )    (Y(I)      )

OBS#
   1        86.00000        88.00000
   2        71.00000        77.00000
   3        77.00000        76.00000
   4        68.00000        64.00000
   5        91.00000        96.00000
   6        72.00000        72.00000
   7        77.00000        65.00000
   8        91.00000        90.00000
   9        70.00000        65.00000
  10        71.00000        80.00000
  11        88.00000        81.00000
  12        87.00000        72.00000


Option number = ?
0                                                        Exit list procedure
SELECT ANY KEY                                           Select special function key labeled-ADV. STAT.
                                                         Remove BSDM media
                                                         Insert General Statistics media
Enter number of desired function:
3                                                        Select two paired sample test

VARIABLE NUMBER FOR X =?
1
VARIABLE NUMBER FOR Y =?
2
*******************************************************************************
                               PAIRED SAMPLE TESTS

VARIABLE FOR X --    X(I)
VARIABLE FOR Y --    Y(I)

*******************************************************************************

Enter desired function:
4                                                        Select sign test


SIGN TEST     SAMPLE SIZE IS     12
---------

    NUMBER OF POSITIVE DIFFERENCES =     7
    (THE 1 POINTS WHERE X(I)=Y(I) ARE EXCLUDED FROM THE TEST)
    NUMBER OF OBSERVATIONS USED = 11

    YIELDS AN APPROX. STD. NOR. DEV. =      .90453       No real differences


Enter desired function:
5                                                        Select Wilcoxon Signed Rank test


WILCOXON SIGNED RANK     SAMPLE SIZE IS     12
--------------------
```

```
SUM OF POSITIVE RANKS = 41.5

(USING RANKS OF X(I)-Y(I) AND EXCLUDING THE      1
 POINTS WHERE X(I)=Y(I))
NUMBER OF OBSERVATIONS USED = 11

YIELDS APPROXIMATE STANDARD NORMAL DEVIATES

     1) WITHOUT CORRECTION FOR CONTINUITY :

      A) NOT COMPENSATING FOR TIED DIFFERENCES :       .75574

      B) CONDITIONAL ON THE EXISTING TIED DIFFERENCES :       .75649

     2) WITH CORRECTION FOR CONTINUITY :

      A) NOT COMPENSATING FOR TIED DIFFERENCES :       .71129

      B) CONDITIONAL ON THE EXISTING TIED DIFFERENCES :       .71199
```

Confirms no differences

```
Enter desired function:
6
```

Select Taha's higher power signed rank test

```
HIGHER POWERED SIGNED RANKS    SAMPLE SIZE IS    12
----------------------------

POWER OF THE RANK (MUST BE 2, 3, 4, OR 5)
2
    POWER OF THE RANK IS 2


    SUM OF POSITIVE RANKS SQUARED = 335.75

    (USING RANKS OF X(I)-Y(I) AND EXCLUDING THE 1
     POINTS WHERE X(I)=Y(I))
    NUMBER OF OBSERVATIONS USED = 11
    YIELDS AN APPROX. STD. NOR. DEV. OF        .8284
    CONDITIONAL ON THE EXISTING TIES AND
    WITHOUT A CORRECTION FOR CONTINUITY
```

Again no difference

```
Enter desired function:
7
```

Select Spearman Rank Correlation

```
SPEARMAN'S RHO    SAMPLE SIZE IS    12
--------------

    SUM OF SQUARED RANK DIFFERENCES = 75

          RHO = .73776
```

Seems to indicate that X & Y are related

```
Enter desired function:
8
```

Select Kendall's Tau test

```
KENDALL'S TAU    SAMPLE SIZE IS    12
-------------

    NUMBER OF CONCORDANT PAIRS = 49
    NUMBER OF DISCORDANT PAIRS = 12

          TAU = .56061
```

Also indicates X & Y are related

```
Enter desired function:
10                                                  Exit two paired sample tests

Enter number of desired function:
6                                                   Return to BSDM
```

Examples on Two Independent Samples

Example 1

The following is an example of a two-sample t-test.

```
*****************************************************************************
*                        DATA MANIPULATION                                  *
*****************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                   Raw data
Mode number = ?
2                                                   On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
ANEXMP2:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES



       ANOTHER EXPAMLE


Data file name: ANEXMP2:INTERNAL

Data type is:   Raw data

Number of observations:    13
Number of variables:        1


Variable names:
   1. MEANS

Subfile name      beginning observation  number of observations
 1. FIRST PART                       1                        6
 2. SEC. PART                        7                        7

SELECT ANY KEY
                                                    Select special function key labeled-LIST
Option number = ?
1                                                   List data


         ANOTHER EXPAMLE

Data type is: Raw data
```

```
                         VARIABLE # 1  (MEANS)
  I        OBS(I)      OBS(I+1)     OBS(I+2)     OBS(I+3)     OBS(I+4)
  1       2.00000      3.00000      4.00000      2.00000      3.00000
  6       4.00000      5.00000      4.00000      2.00000      2.00000
 11       6.00000      3.00000      7.00000
Option number = ?

0
SELECT ANY KEY




Enter number of desired function:
2

VARIABLE NUMBER =?
1
*********************************************************************************

                      TWO INDEPENDENT SAMPLE TESTS

VARIABLE --          MEANS
SUBFILE NUMBER FOR THE 'X' DATA?
1
X SUBFILE --         FIRST PART
SUBFILE NUMBER FOR THE 'Y' DATA?
2
Y SUBFILE --         SEC. PART

*********************************************************************************

Enter desired function:
1


TWO SAMPLE t TEST
-----------------

SAMPLE 1
     N =    6
     MEAN =                      3.000000
     VARIANCE =                   .800000
     COEFF. OF VARIANCE =       29.814240
     STD. DEV. =                  .894427

SAMPLE 2
     N =    7
     MEAN =                      4.142857
     VARIANCE =                  3.809524
     COEFF. OF VARIANCE =       47.112417
     STD. DEV. =                 1.951800

t=      1.3147   WITH    DF=    11
PROB (t >      1.3147) =.10769

Enter desired function:
8

Enter number of desired function:
6
```

Exit list procedure

Select special function key labeled-ADV. STAT

Remove BSDM media

Insert General Statistics media

Select two independent sample test

Select two sample t-test

Exit two sample tests

Return to BSDM

Example 2

A cloud seeding experiment was performed using 16 nonseeded and 10 nonseeded days. The amount of rainfall, in inches, was recorded for the seeded (X) and nonseeded (Y) cases.

Three tests to see if the median rainfall was identical were performed, none of which indicates that the two medians differ significantly.

Taha's squared rank test was performed, since it was assumed that greater precipitation amounts are more important, and should therefore be weighted more heavily in this type of experiment.

```
********************************************************************************
*                              DATA MANIPULATION                               *
********************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                              Raw data
Mode number = ?
2                                                              On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
CLOUD:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES



             CLOUD


Data file name: CLOUD:INTERNAL

Data type is:   Raw data

Number of observations:    26
Number of variables:        1


Variable names:
   1. DAYS

Subfile name      beginning observation  number of observations
 1. SEEDED                           1                       10
 2. NONSEEDED          11                       16

SELECT ANY KEY
                                                               Select special function key labeled-LIST


Option number = ?
1                                                              List all data


               CLOUD
```

```
Data type is: Raw data


                          VARIABLE # 1  (DAYS)
  I       OBS(I)      OBS(I+1)     OBS(I+2)     OBS(I+3)     OBS(I+4)
  1       .05000      .72000       .69000       .09000       .04000
  6       .62000      .37000       .23000      1.18000       .26000
 11       .18000      .88000       .12000       .74000       .43000
 16       .10000      .65000       .06000       .09000       .41000
 21       .12000      .41000       .05000       .03000       .32000
 26       .05000
Option number = ?
0
SELECT ANY KEY
```

Select special function key labeled-ADV. STAT

Remove BSDM media
Insert General Statistics

```
Enter number of desired function:
2
```

Select 2 independent sample test

```
VARIABLE NUMBER =?
1
********************************************************************************

                    TWO INDEPENDENT SAMPLE TESTS

VARIABLE --          DAYS
SUBFILE NUMBER FOR THE 'X' DATA?
1
X SUBFILE --         SEEDED
SUBFILE NUMBER FOR THE 'Y' DATA?
2
Y SUBFILE --         NONSEEDED

********************************************************************************

Enter desired function:
2
```

Select median test

```
MEDIAN TESTS
------------

DO YOU WANT THE COMBINED RANKS PRINTED?
YES
                    COMBINED RANKS
    I           FOR X(I)           FOR Y(I)

    1            4.0000            12.0000
    2           23.0000            25.0000
    3           22.0000            10.5000
    4            7.5000            24.0000
    5            2.0000            19.0000
    6           20.0000             9.0000
    7           16.0000            21.0000
    8           13.0000             6.0000
    9           26.0000             7.5000
   10           14.0000            17.5000
   11                              10.5000
   12                              17.5000
   13                               4.0000
   14                               1.0000
   15                              15.0000
   16                               4.0000
```

Both data sets are combined and then ranked from smallest to largest. Tied ranks are assigned to identical data values.

```
I)      TEST STATISTIC, T = 2
       YIELDS A STD. NOR. DEV. OF           .2894
        CONDITIONAL ON THE 5 EXISTING TIES
```

Useful for large samples. Since the values are small do not reject hypothesis of no differences between X and Y.

```
  - - - - - - - - - - - - - - - - - - - - - - - - -

II) CONTINGENCY TABLE ANALYSIS
                            X              Y             TOTAL

                       *****************************
                       *             *             *
    # OF OBS. >        *      6      *      7      *      13
     GRAND MEDIAN      *             *             *
                       *****************************
                       *             *             *
    # OF OBS. <=       *      4      *      9      *      13
     GRAND MEDIAN      *             *             *
                       *****************************

    TOTAL                     10            16            26

     1) YIELDS AN APPROXIMATE CHI-SQUARE VALUE WITH 1 DF OF

          A) USING YATES' CORRECTION FOR CONTINUITY :
                          .16250

          B) WITHOUT CORRECTION FOR CONTINUITY :
                          .65000

     2) FISHER'S EXACT PROBABILITY OF THE
        EXISTING CELL FREQUENCIES OR WORSE :

                          .34408
```

All three values for the two by two table conclude no difference between X' and Y's for middle value.

```
Enter desired function:
3
```

Select Mann-Whitney test

```
MANN-WHITNEY TEST
-----------------

DO YOU WANT THE COMBINED RANKS PRINTED?
NO
    SUM OF THE RANKS OF X = 147.5

     YIELDS AN APPROX. STD. NOR. DEV. OF :           .6583
         CONDITIONAL ON THE 5 EXISTING TIES
```

Designed to see if X's differ from Y's.

Conclude, they do not. For large sample sizes.

```
Enter desired function:
4
```

Select Taha's squared rank

```
TAHA'S SQUARED RANK
-------------------

DO YOU WANT THE COMBINED RANKS PRINTED?
NO
    SUM OF X RANKS SQUARED = 2786.25

     YIELDS AN APPROX. STD. NOR. DEV. OF :           .7605
         CONDITIONAL ON THE  5  EXISTING TIES
```

Useful to see if X's differ from Y's in spread of data sets.

Conclude they do not.

```
Enter desired function:
8
```

Exit from two independent sample tests

```
Enter number of desired function:
6                                                       Return to BSDM
```

Example 3

An investigator is interested in whether there is a significant difference in the time required to pace himself for one mile between a near sea level location and a high altitude location.

Forty five low altitude observations (Y) and forty high altitude observations (X) were collected. It was decided to test whether the two populations from which the investigator sampled have the same distribution.

Both the Cramer-Von Mises and Kolmogorov-Smirnov tests were performed, neither of which indicates that there is a significant difference between low altitude and high altitude pacing.

```
******************************************************************************
*                              DATA MANIPULATION                             *
******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                       Raw data
Mode number = ?
2                                                       On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
ALTITUDE:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


            ALTITUDE


Data file name: ALTITUDE:INTERNAL

Data type is:   Raw data

Number of observations:     85
Number of variables:         1


Variable names:
   1. ALTITUDE

Subfile name      beginning observation  number of observations
 1. HIGH                             1                       40
 2. LOW                             41                       45

SELECT ANY KEY
                                                        Select special function key labeled-LIST
Option number = ?
1                                                       List all data
```

ALTITUDE

Data type is: Raw data

VARIABLE # 1 (ALTITUDE)

I	OBS(I)	OBS(I+1)	OBS(I+2)	OBS(I+3)	OBS(I+4)
1	405.00000	387.00000	400.00000	392.00000	343.00000
6	394.00000	366.00000	389.00000	356.00000	380.00000
11	394.00000	379.00000	359.00000	357.00000	342.00000
16	367.00000	380.00000	395.00000	442.00000	368.00000
21	361.00000	361.00000	360.00000	353.00000	361.00000
26	387.00000	352.00000	385.00000	349.00000	384.00000
31	351.00000	367.00000	364.00000	363.00000	345.00000
36	348.00000	360.00000	353.00000	355.00000	353.00000
41	361.00000	362.00000	359.00000	382.00000	350.00000
46	392.00000	371.00000	398.00000	400.00000	367.00000
51	379.00000	370.00000	365.00000	362.00000	355.00000
56	376.00000	371.00000	369.00000	375.00000	366.00000
61	373.00000	360.00000	374.00000	412.00000	397.00000
66	360.00000	364.00000	377.00000	360.00000	450.00000
71	438.00000	408.00000	380.00000	414.00000	383.00000
76	386.00000	362.00000	380.00000	377.00000	360.00000
81	357.00000	393.00000	357.00000	369.00000	373.00000

Option number = ?

0 — Exit list procedure

SELECT ANY KEY — Select special function key labeled-ADV. STAT

Remove BSDM media

Insert General Statistics

Enter number of desired function:

2 — Select two independent sample test

VARIABLE NUMBER =?

1

****

TWO INDEPENDENT SAMPLE TESTS

VARIABLE -- ALTITUDE

SUBFILE NUMBER FOR THE 'X' DATA?

1

X SUBFILE -- HIGH

SUBFILE NUMBER FOR THE 'Y' DATA?

2

Y SUBFILE -- LOW

****

Enter desired function:

5 — Select Cramer-Von Mises

CRAMER-VON MISES — Hypothesis is that x distribution is the same as y

----

SUM OF THE SQUARED DIFFERENCES .9471

YIELDS A TEST STATISTIC, T= .2359

CRITICAL REGION OF SIZE 0.10 IS FOR T > 0.347

0.05 IS FOR T > 0.461

0.01 IS FOR T > 0.743 — Accept hypothesis

Enter desired function:

6 — Select Kolmogorov-Smirnov test

```
KOLMOGOROV-SMIRNOV
------------------                                      Same hypothesis

     MAXIMUM DIFFERENCE, T (IN ABS. VALUE) =            .2556

     LARGE SAMPLE CRITICAL REGION OF SIZE 0.10 IS FOR T >         .2651
                                          0.05 IS FOR T >         .2955
                                          0.01 IS FOR T >         .3542
                                                        Same conclusion
Enter desired function:
8                                                       Exit

Enter number of desired function:
6                                                       Return to BSDM
```

Example On Multiple Sample Data Sets

1. The following example was run to determine the effect of the addition of different sugars on length (in ocular units) of pea sections grown in tissue culture with auxin present. The first sample contains the control results, while the other samples contain:

   a. 2% glucose added
   b. 2% fructose added
   c. 1% glucose and 1% fructose added, and
   d. 2% sucrose added.

   After running the one way AOV, a large F value was calculated, indicating there was some difference. To determine which samples were different, two multiple comparison tests were run. In both the Least Significant Difference and in the Duncan's test, all samples differed significantly from the control sample. The Kruskal-Wallis test further supports this conclusion.

```
*********************************************************************************
*                               DATA MANIPULATION                               *
*********************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                       Raw data
Mode number = ?
2                                                       On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
TISSUE:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


                        TISSUE CULTURE GROWTH


Data file name: TISSUE:INTERNAL
```

```
Data type is:   Raw data

Number of observations:    50
Number of variables:        1


Variable names:
   1. GROWTH

Subfile name      beginning observation  number of observations
 1. CONTROL                          1                        10
 2. 2% GLUCOSE                      11                        10
 3. 2% FRUCT.                       21                        10
 4. 1%GLU+1FRU                      31                        10
 5. 2%SUCROSE                       41                        10

SELECT ANY KEY
```

Select special function key labeled-LIST

```
Option number = ?
1
```

List all data

```
                          TISSUE CULTURE GROWTH

Data type is: Raw data


                              VARIABLE # 1  (GROWTH)
   I        OBS(I)       OBS(I+1)       OBS(I+2)       OBS(I+3)       OBS(I+4)
   1      75.00000       67.00000       70.00000       75.00000       65.00000
   6      71.00000       67.00000       67.00000       76.00000       68.00000
  11      57.00000       58.00000       60.00000       59.00000       62.00000
  16      60.00000       60.00000       57.00000       59.00000       61.00000
  21      58.00000       61.00000       56.00000       58.00000       57.00000
  26      56.00000       61.00000       60.00000       57.00000       58.00000
  31      58.00000       59.00000       58.00000       61.00000       57.00000
  36      56.00000       58.00000       57.00000       57.00000       59.00000
  41      62.00000       66.00000       65.00000       63.00000       64.00000
  46      62.00000       65.00000       65.00000       62.00000       67.00000
Option number = ?
0
```

Exit list procedure

```
SELECT ANY KEY
```

Select special function key labeled-ADV. STAT
Remove BSDM media
Insert General Statistics

```
Enter number of desired function:
4
```

Select three or more samples

```
NUMBER OF TREATMENTS =?
5

*******************************************************************************

                           MULTIPLE SAMPLE TESTS

VARIABLE --               GROWTH
SUBFILE NUMBER FOR TREATMENT # 1   =
?
1
TREATMENT # 1SUBFILE -- CONTROL
```

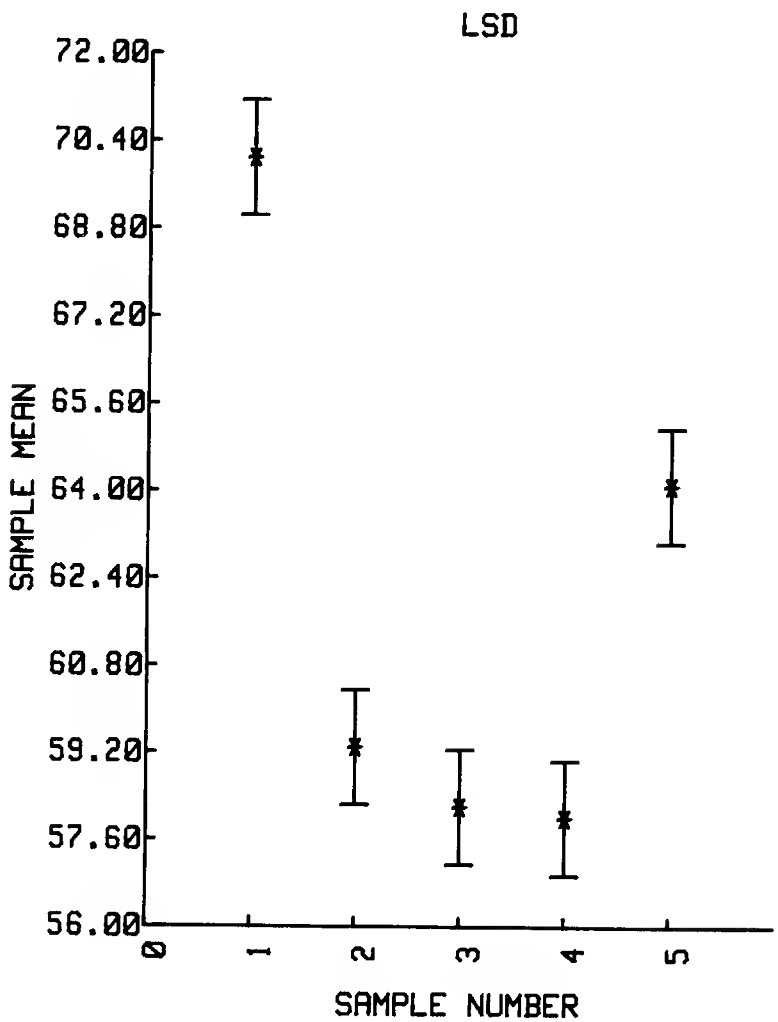
Specify treatments by subfiles

```
SUBFILE NUMBER FOR TREATMENT # 2   =
?
2
TREATMENT # 2SUBFILE -- 2% GLUCOSE
SUBFILE NUMBER FOR TREATMENT # 3   =
?
3
```

```
TREATMENT # 3SUBFILE -- 2% FRUCT.
SUBFILE NUMBER FOR TREATMENT # 4  =
?
4
TREATMENT # 4SUBFILE -- 1%GLU+1FRU
SUBFILE NUMBER FOR TREATMENT # 5  =
?
5
TREATMENT # 5SUBFILE -- 2%SUCROSE

********************************************************************************

Enter desired function:
1                                                     Select one-way AOV

ONE WAY AOV
-----------

     TRT # 1

        75.00000        67.00000        70.00000        75.00000
        65.00000        71.00000        67.00000        67.00000
        76.00000        68.00000

     TRT # 2

        57.00000        58.00000        60.00000        59.00000
        62.00000        60.00000        60.00000        57.00000
        59.00000        61.00000

     TRT # 3

        58.00000        61.00000        56.00000        58.00000
        57.00000        56.00000        61.00000        60.00000
        57.00000        58.00000

     TRT # 4

        58.00000        59.00000        58.00000        61.00000
        57.00000        56.00000        58.00000        57.00000
        57.00000        59.00000

     TRT # 5

        62.00000        66.00000        65.00000        63.00000
        64.00000        62.00000        65.00000        65.00000
        62.00000        67.00000


    TRT.#    N          MEAN          VARIANCE    STD DEV   STD ERRORS

       1     10         70.1000        15.8778     3.9847      1.2601
       2     10         59.3000         2.6778     1.6364       .5175
       3     10         58.2000         3.5111     1.8738       .5925
       4     10         58.0000         2.0000     1.4142       .4472
       5     10         64.1000         3.2111     1.7920       .5667
```

```
                         ANALYSIS OF VARIANCE

      SOURCE      DF           SS                MS                 F


      TOTAL       49         1322.8200
      TRTS         4         1077.3200         269.3300          49.3680
      ERROR       45          245.5000           5.4556

      PROB (F >   49.3680) =0.0000                          Treatments differ significantly

     BARTLETT'S TEST
      DF =  4 ,CHI-SQUARE = 13.9386
           PROB (CHI-SQUARE > 13.9386) = .0075              Variances within treatments also differ.
                                                            Probably just first treatment differs from the others.

Enter desired function:
2                                                           Select multiple comparisons


MULTIPLE COMPARISONS
--------------------

CHOOSE A NUMBER AND PRESS CONTINUE
1
WHAT CONFIDENCE LEVEL ? (.99,.95,etc.)
.95                                                         LSD procedure at 95% confidence.
TABLE VALUE FROM STUDENT'S t
2.02
DO YOU WISH TO PLOT ON THE CRT?
YES
Beep signify the end of plot, then press CONTINUE.
DO YOU WANT A HARD COPY(IF THIS IS FEASIBLE)?
NO

    LSD

         ERROR MEAN SQUARE =     5.4556
         DEGREES OF FREEDOM = 45
         CONFIDENCE LEVEL = .95
         TABLE VALUE FROM STUDENT'S t =     2.0200, LSD =     2.1100


       SAMPLES RANKED


           4  3  2  5  1
         A  -------
         B           -
         C              -

       MEANS
          1 -C                                              Treatments 2-4 are not different from one another.
          2 -A
          3 -A                                              Treatment 1 differs from the others.
          4 -A
          5 -B                                              Treatment 5 differs from the others.
CHOOSE A NUMBER AND PRESS CONTINUE
1
WHAT CONFIDENCE LEVEL ? (.99,.95,etc.)
.95
TABLE VALUE FROM STUDENT'S t
2.02
DO YOU WISH TO PLOT ON THE CRT?
NO
Plotter indentifier string(press CONT if'HPGL')?
```

```
Plotter select code, bus #(defults are 7,5)?

WHICH PEN COLOR SHOULD BE USED?
1
Beep signify the end of plot, then press CONTINUE.

    LSD

         ERROR MEAN SQUARE =     5.4556
         DEGREES OF FREEDOM = 45
         CONFIDENCE LEVEL = .95
         TABLE VALUE FROM STUDENT'S t =     2.0200, LSD =     2.1100


      SAMPLES RANKED


          4  3  2  5  1
        A  -------
        B           -
        C              -

      MEANS
         1 -C
         2 -A
         3 -A
         4 -A
         5 -B
```



```
CHOOSE A NUMBER AND PRESS CONTINUE
5                                                       Choose Duncan's multiple comparison procedure
ERROR MEAN SQUARE =?
5
DEGREES OF FREEDOM =?
2
WHAT CONFIDENCE LEVEL ? (.99,.95,etc.)
.95
TABLE VAL FROM NEW MULT RANGE TEST FOR  5 MEANS        Tables available in appendix
?
3.17
TABLE VAL FROM NEW MULT RANGE TEST FOR  4 MEANS
?
3.1
TABLE VAL FROM NEW MULT RANGE TEST FOR  3 MEANS
?
3.01
TABLE VAL FROM NEW MULT RANGE TEST FOR  2 MEANS
?
2.86

    DUNCAN'S TEST

         ERROR MEAN SQUARE =     5.0000
         DEGREES OF FREEDOM = 2
         LEVEL OF CONFIDENCE = .95

     NUMBER OF MEANS = 5, TABLE VALUE =     3.170 , DIFFERENCE =     2.242
     NUMBER OF MEANS = 4, TABLE VALUE =     3.100 , DIFFERENCE =     2.192
     NUMBER OF MEANS = 3, TABLE VALUE =     3.010 , DIFFERENCE =     2.128
     NUMBER OF MEANS = 2, TABLE VALUE =     2.860 , DIFFERENCE =     2.022

       SAMPLES RANKED


           4  3  2  5  1
         A  -------
         B           -
         C              -

       MEANS
          1 -C
          2 -A
          3 -A                                          Same conclusion as in LSD
          4 -A
          5 -B
CHOOSE A NUMBER AND PRESS CONTINUE
6                                                       Exit multiple comparisons

Enter desired function:
3                                                       Choose Kruskal-Wallis test


KRUSKAL-WALLIS TEST
-------------------

     CHI-SQUARE =     38.1101       DF =   4            Conclude treatments differ.
     P(CHI-SQUARE )     38.1101) = 0.0000

Enter desired function:
5                                                       Exit 3 or more samples

Enter number of desired function:
6                                                       Return to BSDM
```

Analysis of Variance

General Information

Description

The Analysis of Variance package is made up of six analysis routines as well as a number of auxiliary routines that can be used after the analysis of variance (ANOVA or AOV) is completed.

The following analyses are available for balanced data sets –

- **Factorial design** - multiway classification with or without major blocks.
- **Nested design** - includes completely nested, mixed nested and crossed classifications.
- **Split-plot design** - several types in which one or more factors can be in the whole plot.

These three analyses can be used for balanced or unbalanced designs -

- **One-way ANOVA** - completely randomized one-way classification.
- **Two-way ANOVA** (unbalanced) - one or more of the cells can be empty or be unequal in sample size.
- **One-way Analysis of Covariance** - for the completely randomized one-way classification.

For each of the designs in this package, the objective of the routine is to sort out the sources of variability and assign, if possible, responsibility for a portion of the total variability in the data to certain factors in the design.

Input

The first step is to input your data via the Basic Statistics and Data Manipulation routines. Because the data for the AOV programs must be in a very structured format, please read the Basic Statistics and Data Manipulation section of this manual and the portion of this section entitled Data Structures before entering your data. After entering your data, one of the six types of designs is selected and questions will be asked in order to determine the exact design you are using.

Auxiliary Routines

The following routines can be used to complement the analyses performed by the six design routines -

- **Orthogonal Polynomials** - performs a decomposition of the specified sum of squares into linear, quadratic,...,portions. This routine should be used only for factors with quantitative levels.
- **Treatment Contrasts** - performs a comparison on a specified factor. Output includes sum of squares and F ratio.
- **Multiple Comparison Procedures** - can be used to perform one or more of five routines to determine which factor levels represent different population levels. For a more detailed description, please see the portion of this manual entitled Multiple Sample Tests in the General Statistics section.
- **Interaction Plot** - allows you to study the relationship between two or three factors. (Not available from One-way or Covariance routines.)
- **FPROB** - generates right-tailed probability values for the F distribution.

Special Routines

New Response

This allows you to specify a new response variable for the last design chosen. So, even after you have done multiple comparisons (or any other analysis) you may go back to the same design and specify a new response variable without having to answer all of the design questions.

After this is done, a title and description of the last design will be displayed on the CRT.

Special Considerations

Limitations

This program is capable of handling 50 variables with a total of 1500 data values. In addition, there are certain limitations imposed for each program as follows -

- **Factorial** - the product of (levels of A)*(levels of B)*(levels of C)*(levels of D) = size ≤ 500. Also, (number of blocks)*size*(number of observations per cell) ≤ 1500.
- **Nested** - size (as described above) ≤ 500. No blocks are permitted.
- **Split Plot** - Blocks are necessary. Only factors A,B and C are permitted in addition to blocks, and (levels of A)*(levels of B)*(levels of C)*(number of blocks) ≤ 500.
- **One Way** - There can be up to 50 treatments.
- **Two Way (unbalanced)** - At least one cell must have more than one observation. The number of rows (A factor) ≤ 20. The number of columns (B factor) ≤ 20. (number of rows)*(number of columns) ≤ 200.

- **One-way Covariance** - There can be up to 25 treatments.
- **Orthogonal Polynomial** - The polynomial can be up to the tenth degree.
- **Treatment Contrast** - There can be up to 20 levels of one-way means and up to 200 levels of two-way means.
- **Multiple Comparison** - same as for Treatment Contrast.
- **Interaction Plot** - there can be no more than 20 levels of the factor plotted on the X axis, otherwise the plot becomes "messy".

Balanced vs. Unbalanced Designs
To convert from a balanced design to an unbalanced design, you need to use the data manipulation section of the package to create variable(s) with the factor levels for the two factors in the unbalanced design.

On the other hand, if you have finished a factorial analysis and now want to use a one-way design on the same data set, the program allows you to do this by selecting the Advanced Statistics option on the menu.

Discussion

General

The analysis of variance (AOV) technique can be used in many data analysis situations where it is desired to characterize the sources of variation in a "planned" experiment. The essential feature of AOV is that the total variation of the numbers (data) is uniquely decomposed into separate parts. For example, suppose we have run an experiment in which we used four varieties of corn and three row spacings. We repeated this experimental set-up five times (on five fields). We can then break the total variation down into five components as indicated below:

AOV

Source	DF	SS	MS	F
Total	$5*4*3-1=59$	SS_T		
Fields (or Blocks)	$5-1=4$	SS_B	MS_B	$F_1=MS_B/MS_E$
Varieties	$4-1=3$	SS_V	MS_V	$F_2=MS_V/MS_E$
Row Spacings	$3-1=2$	SS_R	MS_R	$F_3=MS_R/MS_E$
Var. X Row	$3*2=6$	SS_{VR}	MS_{VR}	$F_4=MS_{VR}/MS_E$
Error	44	SS_E	MS_E	

In order to more fully develop our understanding of the usefulness of AOV, let us discuss how one might use such a table. Starting with the first column, we see the decomposition of the total variation into its five components. The next column shows the allocation of the so-called degrees of freedom (see references). Notice that the degrees of freedom components add up to the degrees of freedom associated with the total sum of squares. For the total source of variation, the degrees of freedom will be the total number of observations in the experiment minus one. The SS(sum of squares) column shows the breakdown of the total sum of squares for the experiment into the various components. One could prove algebraically that $SS_T = SS_B + SS_V + SS_R + SS_{VR} + SS_E$ and likewise for the degrees of freedom. The MS (mean square) column is obtained by taking SS/DF. This reflects an "average" variation due to each of the sources.

The last column is the F-ratio or testing column. Generally, we are testing the hypothesis that there is "nothing" happening in the experiment versus the expected hypothesis that something "worthwhile" is occurring. If nothing is happening, then all mean sources of variation should be of the same magnitude as the error mean square. The F-ratio is a statistical test to see if the mean square for the source of variation in question is significantly bigger than the error mean square. If it is, we can conclude that there is a "real" effect. For example, suppose that F_2 is quite large. We would then be able to conclude that the population variety means are not all the same. That is, at least one of the variety means differs significantly from the others.

How big do the F values have to be? That depends on the degrees of freedom associated with the numerator MS and the degrees of freedom associated with the denominator (error) MS. The computed F values may be compared with tabled values to find out if they are significant at the .10, .05, .01, or .005 level, or, with this program, you can actually compute the level of significance. The program will automatically calculate the Prob[F > F calculated] for a factorial AOV. For nested or partially nested AOV, the user may elect to use the F probability option to find the probability levels.

Factorial Versus Nested Models

Many researchers have difficulty differentiating between a factorial model and a nested model for AOV. A brief example may be of some help. In a three-way factorial model, for example, the levels of factor B are the same over all levels of factors A and C. Suppose factor A is three temperature settings, factor B is two pressure settings and factor C is four different laboratories. In a factorial model, we would assume that each of the six (three temperature * two pressure) combinations had been studied at each of the four laboratories. In a nested AOV with factor C nested in A and B, we might assume that the same six combinations were run; however, for each of the six combinations, four different laboratories (greenhouses, plants, fields, classrooms, etc.) were used. Hence, a total of 24 laboratories were used instead of just four. Assuming just one observation per laboratory and experimental combination, the AOV table for the factorial would be:

Factorial AOV Example

Source	DF	SS	MS
Total	3*2*4 − 1 = 23	SS_{Total}	
Temperature	3 − 1 = 2	SS_T	MS_T
Pressure	2 − 1 = 1	SS_P	MS_P
Temp x Pres	2 * 1 = 2	SS_{TP}	MS_{TP}
Laboratories	4 − 1 = 3	SS_L	MS_L
Temp x Lab	2 * 3 = 6	SS_{TL}	MS_{TL}
Pres x Lab	1 * 3 = 3	SS_{PL}	MS_{PL}
Temp x Pres x Lab	2*1*3 = 6	SS_{TPL}	MS_{TPL}

However, for the nested model described above, the AOV table would be:

Nested AOV Example

Source	DF	SS	MS
Total	23	SS_{Total}	
Temperature	3 − 1 = 2	SS_T	MS_T
Pressure	2 − 1 = 1	SS_P	MS_P
Temp x Pres.	2 * 1 = 2	SS_{TP}	MS_{TP}
Lab (temp x pres)	(4 − 1)*3*2 = 18	$SS_{L(TP)}$	$MS_{L(TP)}$

Notice that the AOV tables are somewhat different. Actually, the $SS_{L(TP)}$ can be obtained (and is in the program) from the first AOV table by noting that $SS_L(TP) = SS_L + SS_{TL} + SS_{PL} + SS_{TPL}$. Generally, in nested or partially nested AOV's, the nested factor is considered to be a random effect.

Partially Nested vs. Nested Models

Consider a laboratory experiment involving mice in which three levels of some drug (factor A) are to be investigated. Seven mice (factor B) are used for each drug level and the response variable is determined on four days (factor C). One model which might be used for the analysis would be three levels of factor A; seven levels of factor B nested on factor A; and four levels of factor C. The AOV table would be:

AOV

Source	DF	SS	MS
Total	83	SS_{Total}	
Drug	2	SS_D	MS_D
Mice(Drug)	18	$SS_{M(D)}$	$MS_{M(D)}$
Days	3	SS_T	MS_T
Drug x Days	6	SS_{DT}	MS_{DT}
Time x Mice(Drug)	54	$SS_{TM(D)}$	$MS_{TM(D)}$

This type of design is sometimes called a repeated measurements design. It is also a partially nested design because factor C is crossed both with factor A and the nested factor B. As is indicated by the arrows in the AOV table, at least two different "error" terms are used for studying the significance in this model. It should be noted that it is necessary to have exactly the same number of subjects within each level of factor A in order to use the analysis in this package.

Two-Factor AOV Structure

The analysis of variance is a method of decomposing the sum of squared deviations of the observations about the overall mean $[\Sigma(y_{ijk} - \bar{y}_{...})^2]$ into various sources. For a two-factor design, we may show sources of variation due to the row effect (A), the column effect (B), the row-by-column interaction effect (AB) and the within error effect (ERROR). For example, consider an experiment in which we have four levels of temperature (100, 150, 175, 200°C) and three levels of pressure (5, 10, 15 psi) with several determinations of the chemical yield (y) for each combination of temperature (ROWS) and pressure (COLUMNS). One possible arrangement of the data might be as shown below:

		Pressure		
		5	10	15
Temperature		**Column 1**	**Column 2**	**Column 3**
100	Row 1	y_{111} y_{11n11}	y_{121} y_{12n12}	y_{131} $.y_{13n13}$
150	Row 2	.	.	.
175	Row 3	.	.	.
200	Row 4	$y_{411}.$ $.y_{41n41}$	$y_{421}.$ $.y_{42n42}$	$y_{431}.....y_{43n43}$

Each y_{ijk} stands for the numerical value of the chemical yield in percent. The subscript i refers to the row designator, the j for the column designator, and the k for the observation number in the i,jth cell. Notice that the n_{ij} are not necessarily all equal, nor is it necessary that n_{ij} be >= 1. If the n_{ij} are all equal, the analysis of variance involves the usual summing and summing of squares, a task which could be performed by hand calculators. When the n_{ij} are not all equal, the exact analysis is quite complicated.

Note that the table which we have described above does not show how the experiment was actually run. According to good statistical practice the order of running the experiment should be in a random fashion. That is, conceptually, all of the possible sequences should be equally likely and the experimenter should choose one sequence at random.

Reasons for Unbalanced Designs

Unbalanced two-factor designs might arise in at least three ways. First, the design could have been planned as a balanced design (all n_{ij} equal). However, several observations may be lost due to death of a subject, etc. This often happens in research even though experimenters use good experimental techniques. Second, because of the nature of the variability of one response (or some other reason), the experimenter may have set up the design with an unequal number of observations in the cells. For example, suppose that one of the row levels is really a control or standard dose. It may be a common practice to use fewer observations on the control than the other drugs (other "levels" of the row factor). A third possibility is that certain combinations of the row and column levels might yield results which are impossible to monitor in an experiment. This might happen if in the experiment described above, the highest temperature level (200°C) and the highest pressure level (15 psi) proved to be "too much" for the chemical process. In general, of course, it is not a good procedure to design two-factor experiments in which certain levels of the factors cannot be included in the experiment.

Approximate Analyses for Two-Factor Experiments

If each cell (row-column combination) has at least one observation and the number of observations in each cell is approximately the same, the method of unweighted means is sometimes used. Essentially, in this analysis, the cell means are subjected to the usual two-way AOV with one observation per cell, and the within error term is added to the table after adjustment. (See Bancroft, reference 1, p. 35.) This approximate analysis will probably allow you to draw accurate conclusions for most sets of data.

One reason why we might use this type of analysis is because the "exact" analysis is quite complicated. The complexity of the analysis is related to the fact that the calculations which must be performed do not just involve the usual summing and summing of squared values. In short, the exact analysis is a "messy" problem.

Unbalanced Two-Way AOV - "Exact" Solutions

As described more completely in reference 1, Chapter 1, the solution involves rather messy notation. We shall avoid the notational problems by describing, in words, the procedures that you should use in interpreting the AOV tables, rather than describing the computing procedures which were used.

Once again, the idea of the AOV is to separate out the various sources of variation from an observable set of data. In the balanced two-factor design, the analysis of variance table might be written as follows:

AOV

Source	df	Sum of Squares	Mean Squares
Total	N − 1	TSS	
Rows	R − 1	RSS	RSS ÷ (R − 1)
Columns	C − 1	CSS	CSS ÷ (C − 1)
RxC Interaction	(R − 1)(C − 1)	ISS	ISS ÷ (R − 1) (C − 1)
Residual	N − RC	ESS	ESS ÷ (N − RC)

In this table, R equals the number of rows, C equals the number of columns, and N equals the number of observed y's. The computations which are involved in obtaining the Sum of Squares column will not be described. Suffice it to say that in each case the individual observations or the means are compared to the overall mean.

As a brief review, let us examine that AOV procedure. According to the AOV procedure, we are trying to determine if the source of variation for rows, columns, and/or the interaction is significantly bigger than the error source of variation. This is done by calculating certain ratios of mean squares--the so-called F-ratios. Under the assumption of no differences among the row population means (i.e., levels of temperature), the mean square (MS) for rows should be of the same magnitude as the MS for the error. In a similar fashion, the source of variation for columns and interaction can also be tested.

For balanced sets of data, that is where the subclass frequencies are all the same, the decomposition of the sources of variation for a two-factor design is orthogonal. This means that every SS and MS in the table represents the source of variation as indicated in that row. When we have an unbalanced design, the table is not as easy to interpret.

In order to understand the output provided by this program, we will use the hypothetical experiment described earlier. Suppose that the table of n_{ij}, the frequency counts for the twelve row-column cells is as follows:

Pressure

		5	10	15	
Temperature	100	5	4	5	N = 54
	150	5	5	5	
	175	5	5	4	
	200	4	3	4	

Ordinarily we would ask the investigator to use equal n_{ij}; however, there might be perfectly good reasons why this was not possible.

Preliminary AOV Tables

The next output from this program is the Preliminary AOV tables. The first table has the general form:

Preliminary AOV

Source	DF	SS	MS	F-ratio
Total	$N-1 = 53$	SS_T		
Subclass*	$RC-1 = 11$	SS_S	MS_S	MS_S/MS_E
ERROR	$N-RC = 42$	SS_E	MS_E	

* Rows + Columns + Interaction

The decomposition in this table looks as if we have twelve individual treatments rather than four temperature and three pressure combinations. If the F-ratio is large (and the F-Prob is small), say less than about .05, we can conclude that not all twelve population means are the same. The second table has a further decomposition of the subclass source into main effect differences and interaction differences.

Interaction Preliminary AOV

Source	DF	SS	MS	F-Ratio
Total	$N-1 = 53$	SS_T		
Main Effects*	$R+C-2 = 5$	SS_M	MS_M	MS_M/MS_E
Interaction**	$(R-1)(C-1) = 6$	SS_I	MS_I	MS_I/MS_E
Error	$N-RC = 42$	SS_E	MS_E	

* Row + Column
** RxC

This table helps us determine if there is interaction in our two-way design. This is important because it may help us decide which analysis to use next, that is, which of the FINAL AOV's we should choose (see Bancroft).

If one or more cells are empty, the method of fitting constants must be used for the final analysis. For the method of fitting constants, we assume no interaction is present in the model. Hence, if either one $n_{ij} = 0$ and/or interactions are assumed to be absent in the population, we should use the METHOD OF FITTING CONSTANTS FINAL AOV. If interaction between the row and column factors is expected to be present in the population and all $n_{ij} >= 1$, the METHOD OF SQUARED MEANS should be used.

If you are uncertain whether or not interactions are present, your interpretation of the output of the PRELIMINARY AOV table for interactions may help you decide. If the F-PROB for the interaction F-ratio is small enough, we might conclude that interaction is present. (Bancroft, reference 1, suggests that if F-PROB $< .25$, one should use the method of squared means.)

Interpreting the Method of Fitting Constants AOV

Since this method assumes that the model is of the form Y = A + B * (ROW LEVELS) + C * (COLUMN LEVELS) + ERROR, what remains to be tested by this method is if the row levels (means) differ significantly from each other and if the column levels (means) differ significantly from each other. The calculations involve (see page 16, Bancroft) finding the solution to a set of least-squares equations. As we discussed above, when all n_{ij} are equal, the sum of squares due to rows is orthogonal to the sum of squares for columns. However, when the n_{ij} are not all equal, by using the method of fitting constants, the program will construct the following table:

Source	DF	SS	MS	F-Ratio
Total	$N-1 = 53$	SS_T		
Rows (unadjusted)	$R-1 = 3$	SS_R	MS_R	
Columns (adjusted)	$C-1 = 2$	SS_{C-A}	MS_{C-A}	$F_1 = MS_{C-A}/MS_E$
Columns (unadjusted)	$C-1 = 2$	SS_C	MS_C	
Rows (adjusted)	$R-1 = 3$	SS_{R-A}	MS_{R-A}	$F_2 = MS_{R-A}/MS_E$
Interaction	$(R-1)(C-1) = 6$	SS_I	MS_I	$F_3 = MS_I/MS_E$
Error	$N-RC = 42$	SS_E	MS_E	

The first two F-ratios can be used to test the following hypotheses:

H_o: The "B" terms in the model are not needed; H_o: The "C" terms in the model are not needed. The third F-ratio is the same test for the interaction obtained in the preliminary AOV table. Notice that the SS for columns is obtained after correction for rows. That is, SS_{C-A} (columns adjusted for rows) = SS_M (main effects in preliminary AOV table) - SS_{rows} (rows ignoring the column effects). Hence, some of the calculation for the final AOV by the method of fitting constants are derived from the preliminary AOV table.

In conclusion, the method of fitting constants allows us to make "good" tests for main effects if the interaction term is absent. Also, if one or more n_{ij} = zero we must use this method since the interpretation of a significant interaction is questionable anyway. After determining that the row and/or column means differ significantly, one might wish to do some type of multiple comparison procedure to determine where the significant differences lie.

Interpreting the Method of Squared Means AOV

When interaction is assumed present in our model or suspected to be present in the model after studying the preliminary AOV table, the method of squared means can be used to find "good" estimates of the main effects if all $n_{ij} > 0$. This analysis operates on the cell means weighted by $W_i = c^2/(\sum_j 1/n_{ij})$ for the ith row and $W_j = r^2/(\sum_i 1/n_{ij}$ for the jth column. The model for this situation would be:

$$Y = A + B * (\text{ROW LEVEL}) + C * (\text{COLUMN LEVEL}) +$$

$$D\,(\text{ROW, COLUMN LEVELS}) + \text{ERROR}$$

where A represents the average value and D represents the coefficient for the interaction term. The method, which is described on pages 24-29 of Bancroft, would yield an AOV table as follows:

Source	DF	SS	MS	F-Ratio
Total	$N-1 = 53$			
Rows (weighted)	$R-1 = 3$	$SS_{R\text{-}W}$	$MS_{R\text{-}W}$	$MS_{R\text{-}W}/MS_E$
Columns (weighted)	$C-1 = 2$	$SS_{C\text{-}W}$	$MS_{C\text{-}W}/MS_E$	$MS_{C\text{-}W}/MS_E$
Interaction	$(R-1)(C-1) = 6$	SS_I	MS_I	MS_I/MS_E
Error	$N-RC = 42$	SS_E	MS_E	

The F-ratios for rows and columns using the weighted cell means will indicate if the main effects are significant. Of course, if the interaction term is already determined to be significant, the interpretation of the main effects must be given careful consideration. Quite frequently experimenters find it useful to plot the subclass means in order to study the "pattern" for the interaction.

Orthogonal Polynomial Breakdown

If the levels of the row and/or column factors are quantitative, it might be of interest to decompose the sum of squares for these terms into single-degree-of-freedom terms for a polynomial model. For example, suppose that the row levels are quantitative such as the temperature levels which we described above (100, 150, 175, 200°C). Since there are four levels, it is possible to fit up to a third degree polynomial to the row levels. Hence, the SS for rows could be decomposed into orthogonal components for linear, quadratic and cubic terms, each with one degree of freedom. The program will perform the elaborate calculations even if the row or column levels are unequally spaced. (For example, the column levels were given as 5, 10, 15 psi. Instead, they could have been 5, 10, 20 psi with unequal spacings between the levels.)

For further information about these procedures, see references 1 and 2.

References

1. Bancroft, T.A. (1968). Topics in Intermediate Statistical Methods. The Iowa State University Press, Ames, Iowa.
2. Searle, S.R. (1971). Linear Models, John Wiley and Sons.

Data Structures

In order to provide for the analysis of six different types of designs the arrangement of the data must be 'presumed' by the program. The material that follows describes the various arrangements within the Basic Statistics and Data Manipulation (BSDM) routines, which are possible for each design. Please read the section dealing with the design which you are considering before attempting to enter your data.

Further information about the designs considered in this package can be found in the Discussion section and in the references.

Factorial Designs

All data to be analyzed with the Analysis of Variance package is entered into memory via the Basic Statistics and Data Manipulation routines. The order in which the data is entered is very important. In general, sampling replications are entered in order, then factors are varied, then blocks are varied. That is, assuming a four-factor design and no sampling replications, the levels of factor D must vary the most rapidly, followed by the levels of C, B, A, and finally the levels of the blocks. Consider an example in which there are two blocks (major replications), two levels of A and three levels of B. Assume for the moment that we do not have any sampling replication and only one response variable. The structure within the Basic Statistics and Data Manipulation (BSDM) program would use only one variable since it is not necessary to store the levels of the factors and blocks when using the (balanced) Factorial program. The structure for this two-way factorial in two blocks would be:

OBS.#	Response Variable 1	Factor B	Factor A	Blocks
1	Y_{111}	B_1	A_1	Block 1
2	Y_{112}	B_2		
3	Y_{113}	B_3		
4	Y_{121}	B_1	A_2	
5	Y_{122}	B_2		
6	Y_{123}	B_3		
7	Y_{211}	B_1	A_1	Block 2
8	Y_{212}	B_2		
9	Y_{213}	B_3		
10	Y_{221}	B_1	A_2	
11	Y_{222}	B_2		
12	Y_{223}	B_3		

Note

The levels of Factor B vary most rapidly while the blocks vary the slowest. The Y's represent numerical data which is the only information stored in BSDM. The first subscript indicates the block, the second indicates the level of factor A and the third designates the level of factor B.

You should remember that it is absolutely essential that you arrange your data in this form prior to entering the BSDM program. Of course, if you are careful, there are ways around the apparent limitation suggested above. Consider the following data set which has already been entered via the BSDM program:

OBS#	Variable (i)	Factor V	Factor U	Blocks
1	Y_{111}	V_1	U_1	Block 1
2	Y_{121}	V_2		
3	Y_{112}	V_1	U_2	
4	Y_{122}	V_2		
5	Y_{113}	V_1	U_3	
6	Y_{123}	V_2		
7	Y_{211}	V_1	U_1	Block 2
8	Y_{221}	V_2		
9	Y_{212}	V_1	U_2	
10	Y_{222}	V_2		
11	Y_{213}	V_1	U_3	
12	Y_{223}	V_2		

First of all, note that blocks (major replications) must vary the slowest. We can use this data structure in the Factorial program by telling the program that factor A, the factor which varies slowly, is factor U and has three levels; while factor B is our factor V and has two levels. Hence, independent of the implied subscripts, levels and ordering, we have considerable flexibility in specifying the factors. We must only make sure the Factor A is the factor which varies most slowly while Factor B is the factor which varies most rapidly.

So far we have described how the data must be structured for the major replications and factors. We will now describe the two modes of data arrangement which are permissible for the minor replications (samples). If you have only one sample per treatment combination, there will be no difference between the two modes.

The first mode assumes that the response variable resides in only one of the variables specified in BSDM. Hence any minor replications/samples will have to be entered as subsequent observations in BSDM. For example, suppose we have a factorial with two blocks, two levels of factor A, and three levels of factor B, with two replications (samples) per factorial combination. The data structure with three different response variables might appear as follows:

	Variables				Factor		
OBS#	**1 = %Ca**	**2 = %Cu**	**3 = %Fe**	**Sample**	**B**	**A**	**Block**
1	X_{11}	X_{21}	X_{31}	1	B_1	A_1	Block 1
2	X_{12}	X_{22}	X_{32}	2			
3	X_{13}	X_{23}	X_{33}	1	B_2		
4	X_{14}	X_{24}	X_{34}	2			
5	X_{15}	X_{25}	X_{35}	1	B_3		
6	X_{16}	X_{26}	X_{36}	2			
7	X_{17}	X_{27}	X_{37}	1	B_1	A_2	
8	X_{18}	X_{28}	X_{38}	2			
9	X_{19}	X_{29}	X_{39}	1	B_2		
10	X_{110}	X_{210}	X_{310}	2			
11	X_{111}	X_{211}	X_{311}	1	B_3		
12	X_{112}	X_{212}	X_{312}	2			Block 2
.							
.							
.							
24	X_{124}	X_{224}	X_{324}	2	B_3	A_2	

The first mode of replicate/sample storage conserves on the use of variables (see Special Considerations for program limitations); however, it does use more observations.

If you have only one response variable in your experiment it may be more efficient to use the second mode for specifying the sampling replications. This mode assumes that each observation in the BSDM program contains all replication values stored one per variable. Hence, the same design described above would appear as follows (here, the subscripts indicate the levels of factor A and factor B, respectively):

	Variables		Factor	Factor	
OBS.	**1 = Rep 1**	**2 = Rep 2**	**B**	**A**	**Block**
1	X_{11}	X_{21}	B_1	A_1	Block 1
2	X_{12}	X_{22}	B_2		
3	X_{13}	X_{23}	B_3		
4	X_{14}	X_{24}	B_1	A_2	
5	X_{15}	X_{25}	B_2		
6	X_{16}	X_{26}	B_3		
.					
.					
.					

One other example is included without comment. Keep in mind that in our examples we have named the factors A, B, C, and D. As long as your data is arranged in some order with one factor varying the most rapidly within another factor, etc; you can call these factors A, B, C, and D where your factor called A will vary the slowest, etc.

Example (Factorial)--two Blocks, two levels of Factor A, three levels of factor B, two sampling replications:

DATA ENTRY OPTIONS

FORM 1

OBS.#				Variable #1
1	Blk_1	A_1	B_1	Rep_1
2				Rep_2
3			B_2	Rep_1
4				Rep_2
5			B_3	Rep_1
6				Rep_2
7		A_2	B_1	Rep_1
8				Rep_2
9			B_2	Rep_1
10				Rep_2
11			B_3	Rep_1
12				Rep_2
13	Blk_2	A_1	B_1	Rep_1
.				
.				
.				

FORM 2

OBS.#				Variable#1	Variable#2
1	Blk_1	A_1	B_1	Rep_1	Rep_2
2			B_2	Rep_1	Rep_2
3			B_3	Rep_1	Rep_2
4		A_2	B_1	Rep_1	Rep_2
5			B_2	Rep_1	Rep_2
6			B_3	Rep_1	Rep_2
7	Blk_2	A_1	B_1	Rep_1	Rep_2
8			B_2	Rep_1	Rep_2
9			B_3	Rep_1	Rep_2
10		A_2	B_1	Rep_1	Rep_2
11			B_2	Rep_1	Rep_2
12			B_3	Rep_1	Rep_2

The order of the observations must be as shown above to get the correct results. In general, the levels of blocks will vary slower than levels of factor A, B, C, D and replicates within cells vary the fastest.

Nested Design

The form of the data structure for the nested or mixed design is quite similar to that previously described for the Factorial Designs. As far as the program is concerned, the nested design is considered to be in a factorial arrangement. The program will calculate the sum of squares, etc., as if the design were a factorial design and then pool the appropriate terms to form the nested or mixed design which you specified.

As you may have already noted, the design must be balanced. This means that if factor C is nested within factor A and is denoted as C(A), then there must be exactly the same number of levels of factor C within each level of factor A. You may wish to refer to the Discussion section to familiarize yourself with the design arrangements for a nested design as compared to a factorial design.

Perhaps an example of a completely nested design structure would be helpful at this time. Suppose that within each of five sections of land we select two lakes at random. From each lake assume that three random positions in the lake are chosen at which we select two samples. Suppose further that the samples are each divided into two beakers and are analyzed separately. Assume that three responses are measured: Y_1 = Var.1 = ppm lead, Y_2 = Var.2 = ppm zinc, and Y_3 = Var.3 = ppm copper.

In this experiment, we will designate the five land sections as the levels of factor A, the various lakes as levels of factor B, and the position as levels of factor C. Notice that factor B is nested in factor A, and that factor C is nested within factor B. These relationships are commonly denoted by B(A) and C(B) respectively.

For the first form of data arrangement, the two samples per position in the lake will be shown as stored in subsequent observations (down) rather than in an additional variable (across). A dash (–) indicates a numerical value which would be entered in BSDM.

Form 1

Obs#	Var1 = Y_1	Var2 = Y_2	Var3 = Y_3	Sample	Position	Lake	Section
1	-	-	-	1	P_1	L_1	Sec 1
2	-	-	-	2	-	-	-
3	-	-	-	1	P_2	-	-
4	-	-	-	2	-	-	-
5	-	-	-	1	P_3	-	-
6	-	-	-	2	-	-	-
7	-	-	-	1	$P_1 = P_4$*	L_2	-
8	-	-	-	2	-	-	-
9	-	-	-	1	$P_2 = P_5$	-	-
10	-	-	-	2	-	-	-
11	-	-	-	1	$P_3 = P_6$	-	-
12	-	-	-	2	-	-	-
.				.	.	.	.
.				.	.	.	
.				.	.	.	.
60	-	-	-	2	$P_3 = P_{30}$	$L_2 = L_{10}$*	Sec_5

* Within each lake the "first" position P_1 has no relationship with the "first" position in another lake; hence we have a total of thirty different lake positions.

** Since each section has two lakes selected from it, there are a total of ten lakes studied in this project.

The other form of data entry for this nested design would use twice as many variables since each sample would be included as another variable rather than another observation. Hence the last row would look like:

	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
Obs#	Var1 = Y_1	Var2 = Y_1	Var3 = Y_2	Var4 = Y_2	Var5 = Y_3	Var6 = Y_3
30	-	-	-	-	-	-

With a little practice you will find that it is quite easy to structure your data so that the Nested Analysis will correctly recognize your data set.

Mixed designs must be entered via the BSDM routines in a similar manner. Keep in mind that whichever factor you call D must have its levels varying more rapidly than factor C which in turn varies faster than factor B. The levels of factor A will change only after each level of factor B have appeared once.

Note

BLOCKS as described in the Factorial Design are not considered for the Nested Design. That is, you will not be asked any questions concerning blocks (major replications) of this design.

Split-Plot Design

In terms of the data structure in the BSDM routine, it is immaterial whether one is using a Split-Plot Design or a Factorial Design. Both designs are the same in terms of the data arrangement in BSDM. Examples representing the two modes of data arrangement for the minor replications (samples) will be shown below. Consider a split-plot experiment in which the pull-off force necessary to remove boxes from a tape is to be studied (see Hicks pp 219-222, 226). Two complete replications (blocks) of the following experiment were performed. Three long strips of tape with boxes attached were chosen to represent three different methods of attaching the boxes to the strips. A chamber was used to study the effects of three humidity levels (50, 70, and 90%) on the pulling force of three boxes. The experimental procedure called for randomly choosing one of the three humidity levels and adjusting the chamber to maintain that level. Two portions of each of the three strips were placed in the chamber for a specified period of time. The pull-force was then measured for each of the six portions of strip. Subsequently, one of the two remaining levels of humidity was randomly chosen and the process was repeated. Finally, the last level of humidity was maintained in the chamber. Upon completion of the first three humidities times three strips times two samples = 18 measurements, the entire process was repeated again in a random fashion.

The reason that this is a split-plot design and not a factorial is because of the ordering of the measurements of pull force. Since it was not deemed possible to randomly investigate the effects of humidity and strip type on the pull force response, we have a restricted randomization of the split-plot type.

The two forms for specifying the sample replications are shown below. Note how the factor names A and B have been assigned to the factors in this experiment and how that corresponds to the data arrangement as shown. Only one response variable is necessary for this design.

FORM 1

OBS#	Y = pull force Variable 1	Sample	B Humidity	A Strip	Block
1	-	1	50%	S1	B1
2	-	2			
3	-	1	70%		
4	-	2			
5	-	1	90%		
6	-	2			
7	-	1	50%	S2	
8	-	2			
9	-	1	70%		
10	-	2			
11	-	1	90%		
12	-	2			
13	-	1	50%	S3	
14	-	2			
15	-	1	70%		
16	-	2			
17	-	1	90%		
18	-	2			
19	-	1	50%	S1	B2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
36	.	2	.	.	.

In this experiment we would specify two blocks (major replications). Factor A (strips) has three levels, factor B (humidity) has three levels, and there are two samples for mode 1 (all samples wihin the same variable). Later, in the Split-Plot Design program, we would specify that factor B (humidity) is the whole plot while factor A (strips) is the subplot. As the experiment is described above, the humidity factor (B) would be in the whole plot even though it does not vary as fast as the strip factor (A). We could have entered our data in a manner which would have had the levels of humidity varying the slowest. Then we would identify humidity as factor A.

The second mode of sample specification for this example would require two variables, say variable one and variable two.

FORM 2

OBS#	Y = pull force Var 1 = Sample 1	 Var 2 = Sample 2	Humidity	B Strip	A Block
1	-	-	50%	S_1	B_1
2	-	-	70%		
3	-	-	90%		
4	-	-	50%	S_2	
5	-	-	70%		
6	-	-	90%		
7	-	-	50%	S_3	
8	-	-	70%		
9	-	-	90%		
10	-	-	50%	S_1	B_2
11	-	-	70%		
12	-	-	90%		
13	-	-	50%	S_2	
14	-	-	70%		
15	-	-	90%		
16	-	-	50%	S_3	
17	-	-	70%		
18	-	-	90%		

One-Way Design

The one-way design, or one-way classification as it is sometimes called, has three possible forms of data organization or structures in BSDM. These three forms are identical to the forms for the ONE-WAY ANALYSIS OF COVARIANCE except that the covariance analysis will expect both a response variable, Y, and a covariate, X, to be specified while the ONE-WAY DESIGN expects only the response variable Y.

The first mode of data organization for the one-way classification uses t variables in BSDM to specify the t treatments in this design. Consider an experiment in which four types of "mums" were investigated in a greenhouse experiment. Suppose two responses were measured: diameter (Y_1) and plant height (Y_2). The data was collected in two separate years (subfiles) with approximately five pots per variety. One possible organization of this data is as follows:

Mode 1 Example

	Variable Response Treatment/Variety OBS#	1 Y_1 Type 1	2 Y_2	3 Y_1 Type 2	4 Y_2	5 Y_1 Type 3	6 Y_2	7 Y_1 Type 4	8 Y_2
Subfile	1	-	-	-	-	-	-	-	-
1975	2	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-
	4	-	-	MV	MV	-	-	-	-
	5	-	-	MV	MV	MV	-	MV	MV
Subfile	6	-	-	-	-	-	-	-	-
1976	7	-	-	-	-	MV	-	-	-
	8	MV	MV	-	-	-	-	-	-
	9	MV	MV	-	-	-	MV	-	-
	10	MV	MV	-	-	-	-	-	-
	11	MV	MV	MV	MV	-	-	-	-

Here, a dash (–) indicates a numerical value is present, and MV indicates that a missing value is assigned to this position.

Note

The arrangement shown above has provisions for missing values to accommodate the various number of pots per treatment (variety). The two subfiles do not have the same number of pots per treatment. The MV operation must be used to 'square-off' the sample sizes for each variable.

You would tell the program that variables one, three, five, and seven represent the four treatments for the first response (diameter). You would then specify the subfile number. The program would then assume that the sample size is five if subfile one is specified and six if subfile two is specified. If subfiles are to be ignored, then a sample size of 11 would be assumed. Of course all calculations within the program would check for missing values (MV) and delete those values from the calculations. Subsequent to the analysis on the first response, Y_1, you may remain within this subfile and specify another response, say Y_2. Finally, you may select another subfile and/or variables for further analysis.

The second mode for possible data organization within the BSDM structure uses only one variable for each response. Within this response variable, the treatment observations are assumed to be contiguous. You specify the number of observations in each treatment including any missing values. The program assumes that the first observation in the first treatment is observation number one if the first subfile is chosen or subfiles are ignored, or the first observation within the specified subfile. Thereafter, the subfile is partitioned into t nonoverlapping but connected intervals - one corresponding to each treatment. Hence, for the example with four treatments and two response variables, one possible arrangement might be:

Mode 2 EXAMPLE

	OBS#	Variable 1 Y_1	Variable 2 Y_2	Treatment# (Variety#)
SUBFILE 1 1975	1	-	-	1
	2	-	-	
	3	-	-	
	4	-	-	
	5	-	-	
	6	-	-	2
	7	-	-	
	8	-	-	
	9	-	-	3
	10	-	-	
	11	-	-	
	12	-	-	
	13	MV	-	
	14	-	-	4
	15	-	-	
	16	-	-	
	17	-	-	
SUBFILE 2 1976	18	-	-	1
	19	-	-	
	20	-	-	2
	21	-	-	
	22	-	-	
	23	-	-	
	24	-	-	
	25	-	-	3
	26	MV	-	
	27	-	-	
	28	-	MV	
	29	-	-	
	30	-	-	
	31	-	-	4
	32	-	-	
	33	-	-	
	34	-	-	
	35	-	-	
	36	-	-	

Note

The sample sizes for the first subfile of each variable would be five, three, four, and four, respectively. For subfile two, the sample sizes would be two, five, five, and six. Of interest is the comparison between the number of data storage positions needed for the two modes of arrangement. For mode 1, the number of positions required would be 11 observations times 8 variables = 88. For the second mode, the number required is 36 observations times 2 variables = 72. In many cases, if there are several missing values you may conserve available memory locations by using the second mode of arrangement.

The third mode of data entry allows for treatments which are not necessarily connected within one variable. Each treatment is composed of a contiguous set of observations. Since this mode of data arrangement may choose treatment groups throughout the data set, it is not possible or necessary to specify subfiles. The arrangement of the data is similar to the arrangement described for method 2, however it is possible to have "gaps" or "holes" in the data set.

Consider the example described above. Suppose it is desired to compare 1975 variety #2 with the 1976 variety #2 for both responses (Y_1 and Y_2). Please refer to the Mode 2 Example and note that we would need to compare observations 6, 7, and 8 with observations 20, 21, 22, 23, and 24. The first three specified observations are from variety #2 in subfile one which is the 1975 data set and the other five values are from variety #2 in subfile two which is the 1976 data set.

Note that although this mode of data arrangement is quite similar to Mode 2, it does provide for more freedom on the part of the data analyst in terms of which treatments are to be used.

Two-Way (Unbalanced) Design

The unbalanced nature of this design makes it more complicated in terms of the data arrangement. It will not be possible to assume that the order of input is completely specified by factor names such as factor A and factor B. This is because it is possible to have not only different numbers of minor replication (samples) within each treatment combination (levels of factor A and factor B), but also to have one or more cells completely missing. Of course, the absence of certain cells is not a desirable characteristic of any factorial experiment; however, there are certain situations in which missing cells naturally occur.

Therefore it is necessary for the BSDM data structure to provide for proper identification of the row and column levels (factors A and B) as well as the particular sample number within that cell. Two methods of specification are permitted for this type of design. The first "data storage type" assumes that you will use three BSDM variables to specify the response variable and factor levels. One variable will be used to store the particular response to be analyzed at this time. One variable will be used for each of the two factors A and B. It is not necessary to use a variable to specify the sample or observation number; however, you may wish to do so in order to completely identify each observation.

Please note that the levels of factors A and B must be the integers 1, 2,...up to the number of levels of each factor. Hence, if factor A has three levels 70, 80, and 120, you would store these three levels in a variable as 1, 2, and 3 rather than 70, 80, and 120. The purpose of this restriction is to conserve data storage allocation. Within the program you will be able to specify the actual levels of the variables when this is necessary for the computation.

As an example of the first data storage type, suppose you have factors of time and temperature involved in an experiment which is designed to study the effects of these two factors on the yield (Y) of a chemical process. Suppose you had used three time settings of 4, 5, and 7.5 hours and three temperature settings of 110, 115, 120° F. Assume that, for one reason or another, from two to five samples were run at each treatment combination (temperature and time condition). Further, let us assume that at the highest temperature and time condition, it was impossible to finish the experimental process. Thus, we can consider this "cell" as missing. Assume two responses Y_1 and Y_2 were measured on almost all samples. One way to enter this data set in the BSDM program is as follows:

Mode 1 Example

BSDM Variable Number

Obs #	1 Y_1	2 Y_2	3 B Levels	4 A Levels	Sample	A Temp	B Time
1	MV	-	1	1	1	110°	4 hrs.
2	-	-	1	1	2		
3	-	-	1	2	1	115°	
4	-	-	1	2	2		
5	-	-	1	2	3		
6	-	-	1	3	1	120°	
7	-	-	1	3	2		
8	-	-	1	3	3		
9	-	-	1	3	4		
10	-	-	2	1	1	110°	5 hrs.
11	-	-	2	1	2		
12	-	-	2	1	3		
13	-	-	2	2	1	115°	
14	-	-	2	2	2		
15	-	-	2	3	1	120°	
16	-	-	2	3	2		
17	-	-	2	3	3		
18	-	-	2	3	4		
19	-	MV	2	3	5		
20	-	-	3	1	1	110°	7.5 hrs.
21	-	-	3	1	2		
22	-	-	3	1	3		
23	-	-	3	2	1	115°	
24	-	-	3	2	2		
25	-	-	3	2	3		
26	MV	MV	3	3	1	120°	

Notes:

1. Observation number 26 is included to let the program know that the cell with temp = 120, time = 7.5 is missing in both responses.
2. Both observation #1 and #19 have one and only one missing response.
3. Although we have shown the 26 observations in a systematic arrangement, this is not necessary except for your own information.
4. The specification of variable numbers in the analysis will identify which factor it should consider as rows (factor A) and which it should consider as columns (factor B).

The second data storage mode allows you to conserve on variables by using only one variable to identify both row and column levels. The levels are "packed" into four digits as xxyy, where xx identifies the row level and yy identifies the column level. Consider the example described above. Using the packed form of storage we will need to allocate at least three variables in the BSDM routine. One variable is needed for each response and one for the 'packed' row/column identification. You may wish to use another variable to identify the sample numbers or you might wish to use the 'space' after the row/column specification. For example, suppose for the third row and second column you wish to identify the observation by the index 74. The packed version would be 0302.74. The program will use only the first four digits 0302 to identify the row and column numbers. Up to 6 digits may be input after the decimal point for identification purposes.

The example described above may be entered via the BSDM routine as follows (for the first ten and the last three observations):

Mode 2 Example

BSDM Variable Number

Obs #	**1 Y_1**	**2 Y_2**	**3 ID xxyy**	**Obs#**	**A Temp**	**B Time**
1	MV	-	0101	1	110°	4 hrs.
2	-	-	0101	2		
3	-	-	0102	1	115°	
4	-	-	0102	2		
5	-	-	0102	3		
6	-	-	0103	1	120°	
7	-	-	0103	2		
8	-	-	0103	3		
9	-	-	0103	4		
10	-	-	0201	1	110°	
.	-	-	-	-		
.	-	-	-	-		
.	-	-	-	-		
24	-	-	0302	2		
25	-	-	0302	3		
26	MV	MV	0303	1	120°	7.5 hrs

One-Way Covariance

The three forms of data arrangement for the one-way analysis of covariance are the same as the one-way design except that both a response variable (Y) and a covariate (X) must be specified. Hence, for the example previously described for mode 1 of the one-way design you would need to specify 12 variables of the BSDM data set and specify a covariate for each treatment set. If different covariates are to be used with the two response variables, then you would need 16 variables. One possible ordering of these variables and treatments for the ith observation is as follows:

	Type 1			Type 2			Type 3			Type 4		
Variable#	1	2	3	4	5	6	7	8	9	10	11	12
	X	Y_1	Y_2	X	Y_1	Y_2	X	Y_1	Y_2	X	Y_1	Y_2

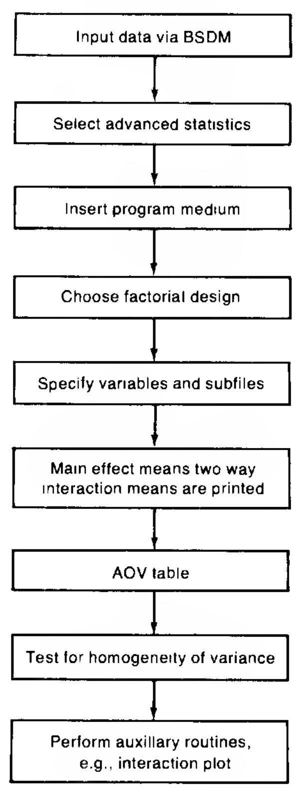
For both mode 2 and 3, you would need to specify one additional variable number as the covariate for each dependent variable. Of course the response variables may use the same covariate in the analysis.

Factorial Design

Object of Program

This program will calculate the complete analysis of variance table for a two-, three-, or four-factor, completely balanced experiment. There may be multiple observations per cell and the entire experiment may be replicated in blocks. The program will automatically print out all main effect and two-way interaction means. If three- or four-way interactions exist, these interaction means may be printed. If there is more than one observation per cell, then tests for homogeneity of variance may be computed. If the experiment has not been replicated, or only one observation per mean is present, there will be no F values computed. All F tests assume that the factors are fixed. A label of up to ten characters may be assigned to each factor.

Typical Program Flow



Special Considerations

See the General Information portion of this AOV manual for program limitations. Also, carefully read the Data Structures section before entering your data through Basic Statistics and Data Manipulation.

References

1. Cochran, W.G. and Cox, G.M., Experimental Designs, John Wiley and Sons, Inc.,1957.
2. Snedecor, G.W. and Cochran, W.G., Statistical Methods, Iowa State University Press, 1967.

Nested or Partially Nested Design

Object of Program

This program will calculate and print the AOV for any valid nested design. The program does this by computing a general factorial and then combining sums of squares to get the desired results. There can be up to five nested factors if samples are entered. This program does not allow the experiment to be replicated in blocks. The program will not compute any F ratios unless the design is a completely nested design. All non-nested main effects, main effect means, and two-way interactions will be printed. If there are any non-nested, three-way interaction means, they may be printed.

Possible Designs

All possible designs are displayed with arbitrary factors P, Q, R and S. In the program you will be asked to match your factors (A, B, etc.) with these arbitrary labels to obtain the design you desire. The notation, Q(P), means that factor Q is nested within factor P. The following options are available.

Number of factors = 2
P
Q(P)

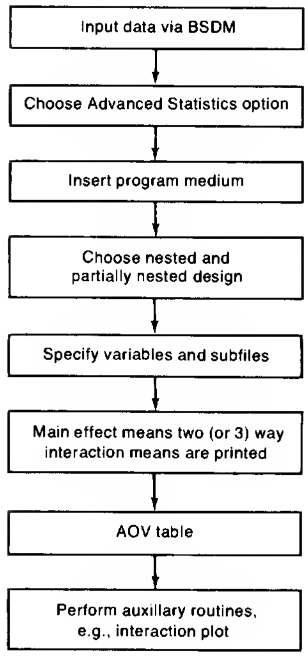
Number of factors = 3

Design 1	Design 2	Design 3
P	P	P
Q(P)	Q	Q(P)
R(Q(P))	PQ	R
	R(PQ)	PR
		QR(P)

Number of factors = 4

Design 1	Design 2	Design 3	Design 4
P	P	P	P
Q(P)	Q	Q	Q(P)
R(Q(P))	R	PQ	R
S(R(Q(P)))	PQ	R(PQ)	PR
	PR	S	QR(P)
	QR	PS	S
	PQR	QS	PS
	S(PQR)	PQS	QS(P)
		RS(PQ)	RS
			PRS
			QRS(P)

Typical Program Flow



Special Considerations

See the General Information portion of this AOV manual for program limitations. Also, carefully read the Data Structure section before entering your data through Basic Statistics and Data Manipulation.

References

1. C.R. Hicks "Fundamental Concepts in the Design of Experiments" 2nd edition. Holt, Rinehart and Winston, 1973.
2. D.C. Montgomery "Design and Analysis of Experiments". Wiley, 1976.

Split Plot Designs

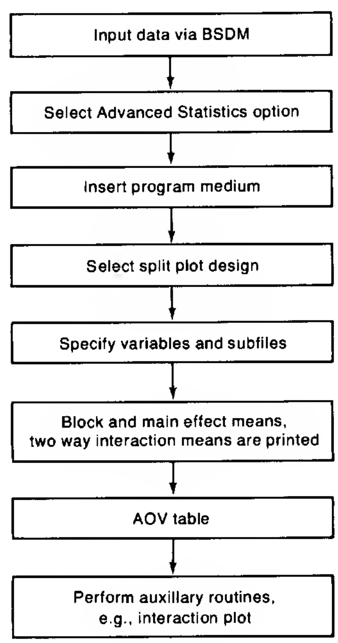
Object of Program

This program will calculate a general factorial and then combine sums of squares to form specific error terms for the split plot or split-split plot design.

Blocks must be present and at least two factors are necessary. Up to three factors may be specified and minor replications (samples) may also be declared.

All main effects and interaction means will be printed. All computed F tests assume the factors are fixed.

Typical Program Flow



Special Considerations

See the General Information portion of this AOV manual for program limitations. Also, carefully read the Data Structures section before entering your data through Basic Statistics and Data Manipulation.

References

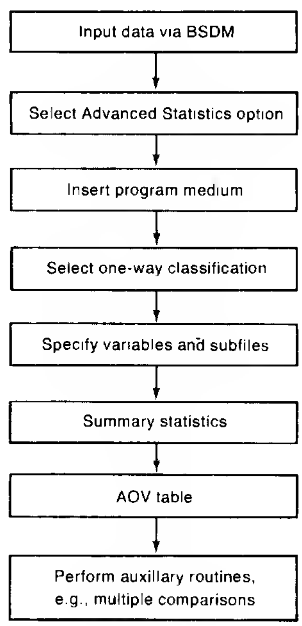
1. C.R. Hicks "Fundamental Concepts in the Design of Experiments" 2nd edition. Holt, Rinehart, Winston, 1973.
2. D.C. Montgomery "Design and Analysis of Experiments". Wiley, 1976.

One-Way Classification

Object of Program

This program will perform a one-way analysis of variance for treatments of equal or unequal size. You may give a ten character name to each treatment. For each treatment the name, sample size, total, mean, and standard deviation will be printed. The analysis of variance table will include all sums of squares and mean squares as well as the calculated F and the probability associated with getting that F value or one larger. You also have control over how many decimal places are to be printed on the output.

Typical Program Flow



Special Considerations

See the General Information portion of this AOV manual for program limitations. Also, carefully read the Data Structure section before entering your data through Basic Statistics and Data Manipulation.

References

1. W.J. Dixon, F.J. Massey "Introduction to Statistical Analysis" Third Edition. McGraw-Hill, 1969.
2. G.W. Snedecor, W.G. Cochran "Statistical Methods" Sixth Edition. Iowa State University Press, 1967.

Two-Way Unbalanced Design

Object of Program

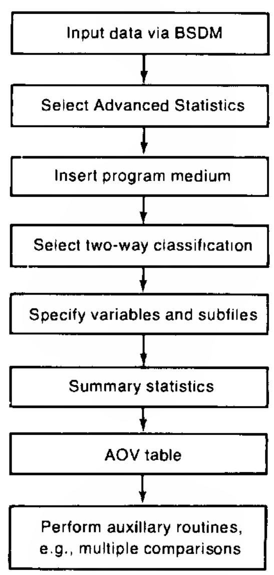
The purpose of this program is to perform an analysis of variance on a two-way classification with unequal subclass frequencies. The analysis may be performed in two ways.

If interactions are known to be present in the population, and all subclasses have at least one observation, then the method of weighted squares of means should be used to test the main effects.

If interactions are known to be absent in the population, or if at least one subclass has no observations, then the method of fitting constants should be used. In any case, if at least one subclass has no observations, the method of fitting constants must be used.

If it is not known whether or not interactions are present in the population, then a preliminary analysis of variance should be studied in order to test for interaction. If this test is significant, then the method of weighted squares of means should be used. A significance level of 0.25 may be used when testing for the presence of interaction.

Typical Program Flow



Special Considerations

See the General Information portion of this AOV manual for program limitations. Also, carefully read the Data Structures section before entering your data through Basic Statistics and Data Manipulation.

References

1. Bancroft, T.A. (1968). Topics in Intermediate Statistical Methods. The Iowa State University Press, Ames, Iowa.
2. Searle, S.R. (1971). Linear Models, John Wiley and Sons.

One-Way Analysis of Covariance

Object of Program

This program will perform a one-way analysis of covariance for equal or unequal sample sizes. You may give a ten-character label to each treatment. For each treatment, a covariate (X) and a response variable (Y) must be specified.

For each treatment, the number of observations in the treatment, the means and standard deviations for the covariate (X) and the response (Y), the correlation between the two, and the equation of the least squares line will be printed. For the overall data, the same things will be computed and printed.

The corrected sums of squares tables will be printed and the analysis of covariance table with the calculated F and the probability associated with getting that F value or one larger will be printed.

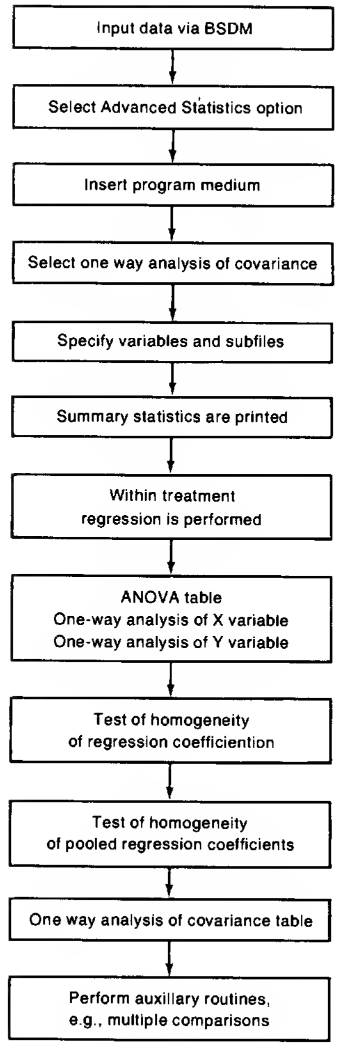
Tests of the one-way analysis of variances for both X and Y, tests for equal slopes within treatments, and significant pooled regression will be calculated and printed.

The adjusted means and the standard errors of the adjusted means will be printed. These adjusted means will be saved for further analysis when doing multiple comparisons, or treatment contrasts.

Any time an observation is found with either the covariate (X) or response (Y) missing, the point will be deleted from the calculations.

You also have control over how many decimal places are to be printed on the output.

Typical Program Flow



Special Considerations

See the General Information portion of this AOV manual for program limitations. Also, carefully read the Data Structure section before entering your data through Basic Statistics and Data Manipulation.

References

1. W.J. Dixon, F.J. Massey "Introduction to Statistical Analysis", Third Edition. McGraw-Hill, 1969.
2. G.W. Snedecar, W.G. Cochran, "Statistical Methods", Sixth Edition. Iowa State University Press, 1967.

F-Prob

Object of Program

Given the numerator degrees of freedom, and the denominator degrees of freedom, and an F value>1, this program will calculate the probability that an F random variable has a value greater than or equal to the given F value.

References

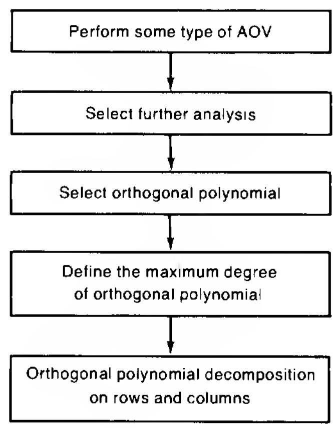
1. Boardman, T.J. (editor) 9830A Statistical Distribution Pac, Hewlett-Packard (PN 09830-70854), September, 1974.
2. Boardman, T.J. (editor) 9845A General Statistics Package.
3. Boardman, T.J. and R.W. Kopitzke, "Probability and Table Values for Statistical Distributions", 1975, Proceedings of the Statistical Computing Section of The American Statistical Association, pp 81-86.

Orthogonal Polynomials

Object of Program

This program generates orthogonal polynomials. This allows you to determine if quantitative factor levels with equal or unequal spacings in the levels are linear, quadratic, etc., in their relationship to the response variable. The output includes the sum of squares, the F-ratio and the P(F>comp F) for each degree polynomial.

Typical Program Flow



Special Considerations

Maximum Degree of Orthogonal Polynomial

For a one-way classification design, it must be less than the number of treatments.

For a two-way (unbalanced) design, it must be less than the number of levels of factor A.

For other designs, it must be less than the number of levels of the factor.

Enter zero if that factor is not a quantitative variable or if it is not desired to do orthogonal polynomial comparisons on the factor.

Level Associated with Treatment (row, factor) #"i"

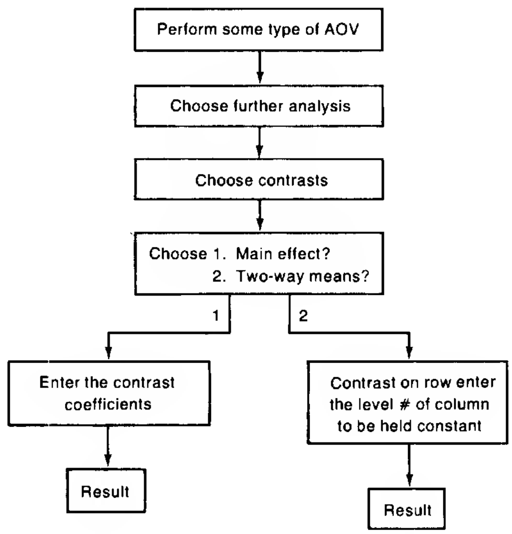
When this question is asked, you should enter the quantity corresponding to this treatment (for one-way design), or this row (two-way design), or the level "i" of factor k (for other design).

Contrasts

Object of Program

This program performs treatment contrasts on main effect means or on two-way means with one of the factors held constant. This allows you to make any desired linear contrast of a set of treatment means by entering an appropriate set of coefficients. The output includes the user-entered coefficients, the contrasts, and the sum of squares, F-ratio and P(F>comp F) associated with the contrasts.

Typical Program Flow



Special Considerations

How to Make a "Contrast"

If the coefficients for the contrasts you enter are denoted by c(i), then one condition for choosing the c(i) is that they must satisfy

$$\Sigma c(i) = 0$$

where i is summed over all levels of the factor of interest. Obviously, this implies that some of the c(i) must be negative. Of course one or more of the c(i) may be equal to zero.

Let's look at an example which demonstrates the procedure. Suppose you have a one-way classification with four treatments. You find in the AOV table that you have a significant F value. So, you reject the hypothesis that all the treatment effects are equal, i.e., you reject

$$H0: T_1 = T_2 = T_3 = T_4.$$

You still don't know exactly which treatments are significantly different from one another. This is where you use a contrast. Suppose you want to know if treatment one is significantly different from treatment three, i.e., you want to test the hypothesis

$$H0: T_1 = T_3, \text{ or } H0: T_1 - T_3 = 0$$

or, written in still another way

$$H0: 1*T_1 + 0*T_2 - 1*T_3 + 0*T_4 = 0$$

If the number of observations in each treatment are equal, then to specify the above contrast all you need to do is to supply the coefficients of the treatments. That is, coefficient one is 1, coefficient two is 0, three is -1 and four is 0. You must tell the program what the coefficients (of the T's above) are.

Suppose the number of observations for the four respective treatments are 6, 8, 7, and 6. Suppose further that you want to test if treatment two is significantly different from treatment four. Write the hypothesis as:

$$H0: 0*T_1 + 1*T_2. + 0*T_3 - 1*T_4 = 0.$$

Then try the following procedure to determine your contrast coefficients, c(i). Form a table using the number of observations for the ith treatment, n(i), as one column. Use the coefficients of the T's in the above hypothesis as the last column. Call these coefficients c(i)n(i).

Remember, one condition for a valid contrast is that $\Sigma c(i)n(i) = 0$. So, check to make sure that condition is satisfied. Then, make a column for your as yet unknown contrast coefficients, c(i). You should have the following table.

n(i)	**c(i)**	**n(i)c(i)**
6		0
8		1
7		0
6		−1

Now, just fill in the c(i) column. To do that notice that $c(i) = n(i)\ c(i)/n(i)$. So you obtain the following.

n(i)	**c(i)**	**n(i)c(i)**
6	0	0
8	1/8	1
7	0	0
6	−1/6	−1

So, contrast coefficient one is 0, two is 1/8, etc.

Notice that the contrast coefficients for a given contrast are not unique. For example, the above contrast would be performed if contrast coefficients of 0, 1/4, 0, −1/3 were given. Also, a similar contrast would be obtained using 0, −1/8, 0, 1/6 as the coefficients.

Interaction Plots

Object of Program

This program will plot two-way interaction, or three-way interaction means. The two-way interaction plot will be on one graph. You may decide which factor will be put on the X axis as well as the spacing of the levels, and then the other factor will be plotted. Each interaction line will be labeled indicating the level of the factor.

For instance, the three levels of a factor B will be labeled B1, B2, B3.

The three-way interaction plot will be plotted on several graphs. That is, a two-way interaction will be plotted for each level of the third factor. The program will give you a prompt when it is necessary to do the next page of the plot.

You may also have a legend drawn showing the length of the Least Significant Difference (LSD) and/or the length of Tukey's Honestly Significant Difference (HSD). To do these, it is necessary to enter the critical value, error mean square, and its corresponding degrees of freedom.

Special Considerations

Which interaction is to be plotted?
When this question is asked, enter the two letters corresponding to the two factors. The input must be one of AB, AC, BC, AD, BD, or CD, and the one selected must be possible for your data set.

What 3-way interaction is to be plotted?
When this question is asked, enter the three letters corresponding to the three factors. The input must be one of ABC, ABD, ACD or BCD.

The label of the X-axis for an interaction plot.
The factor levels must be given in increasing order. Factors whose levels are not in increasing order must be given arbitrary level codes if they are to be used on the X-axis of an interaction plot.

References

1. C.R. Hicks, "Fundamental Concepts in the Design of Experiments"; Second Edition. Holt, Rinehart, and Winston, 1972.
2. B.J. Winer, "Statistical Principles in Experimental Design"; Second Edition. McGraw-Hill, 1971.

Multiple Comparisons

Object of Program

This program allows you to select any one of five multiple comparison procedures to use on either main effect means or two-way table means. You must input the appropriate tabled values for the procedure selected. In addition, for the separation procedures for the two-way means, you will need to specify the appropriate standard deviation to be used.

A separation table will be printed which should help you determine which treatment or factor levels are significantly different from one another. For example, the following table shows output for a set of treatments:

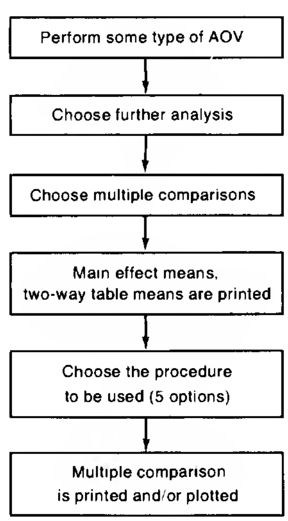
Factor A

Level	Mean	Sample Size	Separation
1	10.7	10	ab
2	9.8	9	a
3	11.7	10	b
4	15.8	8	c

We would interpret this table as showing that factor level 4 is significantly different from the other levels of A since no other level has a "c" listed beside it. Also we see that level 1 cannot be distinguished from level 2 and level 1 cannot be distinguished from level 3. And, level 2 can be shown to significantly differ from level 3 since they have no letters in common.

Of course, the conclusion one draws from the separation procedure may depend on which procedure is used and the level of significance you choose.

Typical Program Flow



Special Considerations

Which factor/main effect should be used?
When this question is asked, you should input A, B, C, or D as the response.

What level of alpha are you going to use?
The value you input in response to this question is used for printout purpose only and not for any calculations.

What table value should you use?
The following chart shows required inputs for tabled values:

Procedure#	Table	Notation	Parameter	Reference
1	Student's t	$t\alpha/2(df)$	df = error degrees of freedom	(1,4)
2	Studentized range	$q\alpha/2(p,df)$	p = # of means df = error degrees of freedom	(6,4)
3	Duncan's	$q^*\alpha(p,df)$	p is as above but reduces by 1 to p=2	(3)
4	Studentized range	$q\alpha/2(p,df)$	p same as 3	(1,3,4)
5	Snedecor's F	$F_2(p-1, df)$	p = # of means	(4,5)

* See references (1) and (2) for more information on all procedures.

Unequal sample sizes
In this case, the harmonic mean, n_0, sample size will be used where $n_0 = p/(1/n_1 + 1/n_2 + \ldots + 1/n_p)$.

For the methods used in Multiple Comparisons, please refer to the Multiple Sample Tests portion of the General Statistics section of this manual.

References

1. Boardman, T.J. and D.R. Moffitt (1971) "Graphical Monte Carlo Type I Error for Multiple Comparison Procedures". Biometrics 27:3, 738-744.
2. Carmen, S.G., and M.R. Swanson (1973) "Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods". Journal of the American Statistical Association 68:341, pp 66-74.
3. Duncan, D.B. (1955). Multiple range and multiple F tests. Biometrics 11, 1-42.
4. Pearson, E.S. and Hartley, H.O. (1958). Biometrika Tables for Statisticians, Vol. I. Cambridge University Press, London.
5. Scheffe,H. (1953). A method for judging all contrasts in the analysis of variance. Biometrika 40,87-104.
6. Tukey, J.W. (1953). The problem of multiple comparisons. Unpublished notes, Princeton University.

Factorial Design

Example

Twenty-four laboratory rats were deprived of food, except for one hour per day, for several weeks. At the end of that time, each rat was inoculated with one of four doses of a certain drug and, after one of three amounts of time, was fed. The weight (in grams) of the food ingested by each rat was measured. The purpose of the experiment is to determine the effect of the drug on the motivation of the rats.

A Time before feeding (hours)	B Dosage (mg/kg) .1	 .3	 .5	 .7
1	9.077	5.63	4.42	1.38
	8.77	8.76	3.01	3.96
5	9.16	11.57	5.22	5.72
	11.82	11.53	9.21	4.69
9	16.08	10.37	7.27	5.48
	14.65	14.46	6.10	9.28

The design for this experiment is a two-way factorial with three levels of time and four dosage levels of the drug. Two rats (observations) per experimental combination were used. The data can be subjected to an analysis of variance in order to determine if there are significant differences between the three times before feeding or the four dosages of the drug. In addition, we can determine if there is a significant interaction between time and dosage.

The F ratios indicate no significant interaction effect ($F = .915$), significant differences in time levels ($F = 14.819$) and dosage levels ($F = 19.533$). The orthogonal polynomial decomposition for the time factor (A) shows a significant linear effect. The decomposition for the dosage factor (B) shows a highly significant linear effect and a cubic effect.

Even though the AB interaction (time or dosage) is not significant, a plot of the two-way means was included to show results of the INTERACTION PLOT routine. A reference LSD value is shown on interaction plot.

```
********************************************************************************
*                              DATA MANIPULATION                              *
********************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                          Raw data
Mode number = ?
2                                                          On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
DEPOFRATS:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES



                      FOOD DEPRIVATION OF RATS


Data file name: DEPOFRATS:INTERNAL

Data type is:   Raw data

Number of observations:    12
Number of variables:        2


Variable names:
   1. OBS 1 WT
   2. OBS 2 WT

Subfiles:  NONE



SELECT ANY KEY
                                                           Select special function key labeled-LIST
Option number = ?
1                                                          List all data
Enter method for listing data:
3

                        FOOD DEPRIVATION OF RATS

Data type is: Raw data


       Variable # 1   Variable # 2
       (OBS 1 WT  )   (OBS 2 WT  )

OBS#
   1         9.07000        8.77000
   2         5.63000        8.76000
   3         4.42000        3.01000
   4         1.38000        3.96000
   5         9.16000       11.82000
   6        11.57000       11.53000
   7         5.22000        9.21000
   8         5.72000        4.69000
   9        16.08000       14.65000
  10        10.37000       14.46000
  11         7.27000        6.10000
  12         5.48000        9.28000
```

```
Option number = ?
0                                                   Exit list procedure
SELECT ANY KEY                                      Select special function key labeled-ADV STAT
                                                    Remove BSDM media
Enter number of desired funtion:                    Insert AOV2
1                                                   Select factorial design
Number of factors in design ? (2, 3, or 4)
2
Number of levels of factor A
?
3                                                   1, 5, and 9 hours
Number of levels of factor B
?
4                                                   .1, .3, .5, and .7 mg/kg
Number of blocks in this design ?
1                                                   Only 1 major replication
No. obs per trt combination in each block(sample)?
2                                                   2 rats per experimental combination
Is the above information correct ?
YES
Do YOU want to assign names to the factors ?
YES
Enter the name for factor A (<11 characters)
?
TIME
Enter the name for factor B (<11 characters)
?
DOSAGE
Data entry option ?
2                                                   Minor replications are stored in different
Variable # for minor replication (sample) 1         variables
?
1
Variable # for minor replication (sample) 2
?
2
No. of decimals for printing calc. values(<=7).
4
********************************************************************************
*                         FACTORIAL ANALYSIS OF VARIANCE                       *
********************************************************************************
                              FOOD DEPRIVATION OF RATS


DESIGN
------
   Number of factors =  2
    No. of levels of factor A =  3
    No. of levels of factor B =  4
    No. of major replications (blocks) =  1
    No. of minor replications (samples) =  2

Subfiles will be ignored
Response variable(s) are :
Variable no.  1      OBS 1 WT
Variable no.  2      OBS 2 WT

MEANS
-----
* Overall mean =          8.2338


* Main Effect Means :

Factor A - TIME   Levels ( 1 - 3 ) :
          5.6250          8.6150          10.4613
```

```
Factor B - DOSAGE    Levels ( 1 - 4 ) :
        11.5917         10.3867          5.8717          5.0850
```

```
* Two Way Interaction Means :

Factor A - TIME  down and  Factor B - DOSAGE  across
                  1              2              3              4
   1            8.9200         7.1950         3.7150         2.6700
   2           10.4900        11.5500         7.2150         5.2050
   3           15.3650        12.4150         6.6850         7.3800
```

```
ANOVA TABLE
-----------
                  Factorial Analysis of Variance

Source (Name)     df     Sums of Squares       Mean Square   F Ratio   F-Prob
-----------------------------------------------------------------------------
Total             23          339.9634           14.7810
A    TIME          2           95.3015           47.6507    14.819    .0006
B    DOSAGE        3          188.4283           62.8094    19.533    .0001
AB                 6           17.6478            2.9413      .915    .5168
Sampling Error    12           38.5858            3.2155
    NOTE: F tests assume that all factors are fixed
```

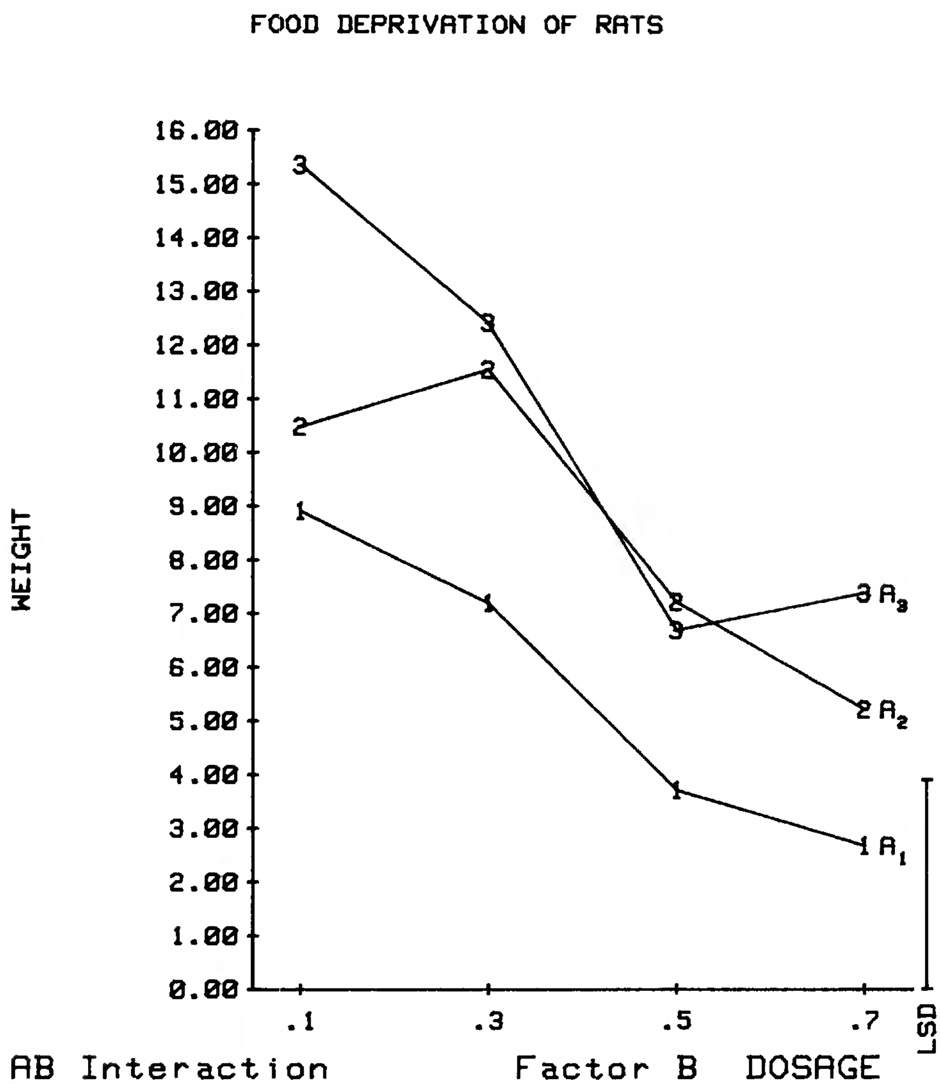
From the AOV table it can be seen that the effects of Factor A and of Factor B are significant, but interaction between Factor A and Factor B is not significant.

```
Should tests for homogeneity of variance be made?
YES
FACTOR LEVELS              CELL STATISTICS

Blk A  B              Mean          Std Dev          Variance   Coef
                                                                Var %
 1  1  1             8.9200          .2121            .0450     2.38
 1  1  2             7.1950         2.2132           4.8984    30.76
 1  1  3             3.7150          .9970            .9941    26.84
 1  1  4             2.6700         1.8243           3.3282    68.33
 1  2  1            10.4900         1.8809           3.5378    17.93
 1  2  2            11.5500          .0283            .0008      .24
 1  2  3             7.2150         2.8214           7.9601    39.10
 1  2  4             5.2050          .7283            .5305    13.99
 1  3  1            15.3650         1.0112           1.0224     6.58
 1  3  2            12.4150         2.8921           8.3640    23.29
 1  3  3             6.6850          .8273            .6844    12.38
 1  3  4             7.3800         2.6870           7.2200    36.41

Bartlett's test :
    Chi squared =  11.0311 with 11 degrees of freedom
    Prob( Chi squared >   11.0311) =  .4410
Specify a new variable for this design ?
NO
Enter desired number:
4
```

Request interaction plot

```
*******************************************************************************
                              INTERACTION PLOT
*******************************************************************************


Is this correct ?
YES                                                         Confirm design on CRT
Plot which factor on the X axis : A,B
?
B
Enter 4 levels of factor B(separate by commas):
?
.1,.3,.5,.7
Name of the response ? ((11 characters)
WEIGHT
Enter Y minimum value. (Less than 2.67 )
?
0
Enter Y maximum value. (Greater than 15.365 )
?
16
Enter Y tic
1
# of decimal places for labelling Y axis((= 6 )=
?
2
Should length of the LSD and/or HSD be plotted ?
YES
Error Mean Square to calculate the LSD and/or HSD.
3.21548                                                     From AOV table
Error Mean Square to be used is 3.21548
t value for the LSD, or 0 not to plot the LSD.
2.179                                                       t-tabled value
Q value for the HSD, or 0 not to plot the HSD.
0
    t = 2.179       LSD = 3.90733040255


Plot on CRT
?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter the select code, bus # (defaults are 7,5)?

Which PEN color should be used?
1
```

```
Beep will sound when plot done, then press CONT.
To interrupt plotting press 'STOP' key
Press CONTINUE when the plotter is ready.
```




```
Are there any more plots to be made ?
NO

Enter number of desired funtion:
9
```

Return to BSDM

Nested or Partially Nested Design

Example

In order to compare two methods of display, a group of six new Thanksgiving greeting cards were selected. Eight stores were selected for the "promotional" display method and another eight stores were used for the "integrated" display method. For each of the two methods and each of eight stores per method, the same six card styles were compared using a response (Y) which measured dollar sales adjusted for store size. The data for each type of display, store, and greeting card style are shown below:

Display Method 1 - "Promotional" (A)

Stores (C)

		1	2	3	4	5	6	7	8
	1	$1.21	1.49	1.76	1.52	0.65	1.96	1.21	1.57
Card	2	1.72	2.09	2.21	2.36	2.83	3.99	2.01	2.62
Style	3	1.72	1.44	1.84	0.91	1.30	7.61	2.01	3.27
(B)	4	0.29	0.92	0.37	0.72	0.43	3.99	2.35	4.71
	5	1.44	2.09	1.84	2.36	1.96	3.26	2.01	1.70
	6	4.43	3.66	0.51	1.78	2.13	5.58	1.41	2.75

Display Method 2 - "Integrated" (A)

Stores (C)

		9	10	11	12	13	14	15	16
	1	$2.60	2.21	1.44	1.20	1.21	3.03	2.79	1.18
Card	2	1.67	1.16	1.73	1.92	4.84	2.88	4.10	1.48
Style	3	3.67	0.78	1.46	1.65	3.23	1.92	4.51	1.48
(B)	4	1.33	0.39	1.33	1.37	2.02	1.68	4.51	2.34
	5	3.33	1.16	1.86	1.92	3.23	2.64	3.96	2.22
	6	4.67	1.90	2.61	3.27	2.26	2.36	2.30	1.55

The mixed nested AOV for this model with factor A (display), factor C (stores) nested in factor A, and factor B (card style) crossed with A and C is shown below. The proper MS for testing differences between the two methods of display is C(A). Notice that the F ratio would be less than one = .42135/4.85529 indicating no significant difference between the methods as well as a considerable amount of store to store variation in the adjusted sales value. There does,however, appear to be significant differences between the population means for card types, i.e. $F = 2.57257/.92726 = 2.77$ which is significant at the .024 level.

A fairly standard procedure for the response variable Y considered here is to transform this response by $Y^* = \ln(Y+1)$ in order to achieve a more homogeneous and consistent response. The next analysis of variance is performed on this new response. The net result is that the F ratio for differences in card type means is even more highly significant (3.93 versus 2.77).

An LSD multiple comparison procedure was done on the six card styles. The results of this comparison show significant differences between style four and all others except style one with certain other differences existing as well. However, if one were looking for the highest adjusted daily sales, one should probably choose one of styles five, two, or six since they were not significantly different from one another but were different from the other styles (although three is questionably different).

```
******************************************************************************
*                              DATA MANIPULATION                             *
******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                      Raw data
Mode number = ?
2                                                      From mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
GRETINGCDS:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES



                THANKSGIVING GREETING CARD EVALUATION


Data file name: GRETINGCDS:INTERNAL

Data type is:   Raw data

Number of observations:    96
Number of variables:        1


Variable names:
   1. DESIGN

Subfiles:  NONE



SELECT ANY KEY
                                                       Select special function key labeled-LIST
Option number = ?
1                                                      List all the data

                  THANKSGIVING GREETING CARD EVALUATION

Data type is: Raw data


                              VARIABLE # 1  (DESIGN)
   I         OBS(I)       OBS(I+1)      OBS(I+2)      OBS(I+3)      OBS(I+4)
   1        1.21000       1.49000       1.76000       1.52000        .65000
   6        1.96000       1.21000       1.57000       1.72000       2.09000
  11        2.21000       2.36000       2.83000       3.99000       2.01000
```

16	2.62000	1.72000	1.44000	1.84000	.91000
21	1.30000	7.61000	2.01000	3.27000	.29000
26	.92000	.37000	.72000	.43000	3.99000
31	2.35000	4.71000	1.44000	2.09000	1.84000
36	2.36000	1.96000	3.26000	2.01000	1.70000
41	4.43000	3.66000	.51000	1.78000	2.13000
46	5.58000	1.41000	2.75000	2.60000	2.21000
51	1.44000	1.20000	1.21000	3.03000	2.79000
56	1.18000	1.67000	1.16000	1.73000	1.92000
61	4.84000	2.88000	4.10000	1.48000	3.67000
66	.78000	1.46000	1.65000	3.23000	1.92000
71	4.51000	1.48000	1.33000	.39000	1.33000
76	1.37000	2.02000	1.68000	4.51000	2.34000
81	3.33000	1.16000	1.86000	1.92000	3.23000
86	2.64000	3.96000	2.22000	4.67000	1.90000
91	2.61000	3.27000	2.26000	2.36000	2.30000
96	1.55000				

```
Option number = ?
0
```

Exit list procedure

```
SELECT ANY KEY
```

Select special function key labeled-ADV STAT
Remove BSDM media
Insert AOV2 media

```
Enter number of desired funtion:
2
```

Choose nested design

```
Number of factors in design ? (2, 3, or 4)
3
Number of levels of factor A
?
2
Number of levels of factor B
?
6
Number of levels of factor C
?
8
Number of samples ?
1
Is the above information correct ?
YES
Which design (by number) is to be used ?
3
```

Shown on CRT, specify design type.

```
Which factor is P: A,B,C
?
A
Which factor is Q: B,C
?
C
Do YOU want to assign names to the factors ?
YES
Enter the name for factor A (<11 characters)
?
DISPLAY
Enter the name for factor B (<11 characters)
?
CARD STYLE
Enter the name for factor C (<11 characters)
?
STORES
No. of decimal places to print calculated values.
4
```

```
********************************************************************************
                         NESTED ANALYSIS OF VARIANCE
********************************************************************************
                     THANKSGIVING GREETING CARD EVALUATION


DESIGN
------
 Number of factors =  3
    No. of levels of factor A =  2
    No. of levels of factor B =  6
    No. of levels of factor C =  8
    No. of minor replications (samples) =  1

Response variable(s) are :
Variable no.  1      DESIGN

MEANS
-----
* Overall mean =          2.2327


* Main Effect Means :

Factor A - DISPLAY   Levels ( 1 - 2 ) :
          2.1665          2.2990
Factor B - CARD STYLE   Levels ( 1 - 6 ) :
          1.6894          2.4756          2.4250          1.7969          2.3112
          2.6981
Factor C - STORES   Levels ( 1 - 8 ) :
          2.3400          1.6075          1.5800          1.7483          2.1742
          3.4083          2.7642          2.2392


* Two Way Interaction Means :

Factor A - DISPLAY  down and  Factor B - CARD STYLE  across
                    1               2               3               4
                    5               6
    1               1.4213          2.4788          2.5125          1.7225
                    2.0825          2.7812
    2               1.9575          2.4725          2.3375          1.8712
                    2.5400          2.6150

Factor A - DISPLAY  down and  Factor C - STORES  across
                    1               2               3               4
                    5               6               7               8
    1               1.8017          1.9483          1.4217          1.6083
                    1.5500          4.3983          1.8333          2.7700
    2               2.8783          1.2667          1.7383          1.8883
                    2.7983          2.4183          3.6950          1.7083

Factor B - CARD STYLE  down and  Factor C - STORES  across
                    1               2               3               4
                    5               6               7               8
    1               1.9050          1.8500          1.6000          1.3600
                     .9300          2.4950          2.0000          1.3750
    2               1.6950          1.6250          1.9700          2.1400
                    3.8350          3.4350          3.0550          2.0500
    3               2.6950          1.1100          1.6500          1.2800
                    2.2650          4.7650          3.2600          2.3750
    4                .8100           .6550           .8500          1.0450
                    1.2250          2.8350          3.4300          3.5250
    5               2.3850          1.6250          1.8500          2.1400
```

	2.5950	2.9500	2.9850	1.9600
6	4.5500	2.7800	1.5600	2.5250
	2.1950	3.9700	1.8550	2.1500

```
Should the 3-way means be printed ?
NO
```

ANOVA TABLE

Nested Analysis of Variance

Source (Name)	df	Sums of Squares	Mean Square
Total	95	148.0541	1.5585
A DISPLAY	1	.4213	.4213
C(A)	14	67.9740	4.8553
B CARD STYLE	5	12.8628	2.5726
AB	5	1.8879	.3776
CB(A)	70	64.9080	.9273

$F = 2.77$ significant at $\alpha = .02$.

There is a significant difference between the population means for card types but not between the types of displays.

```
Enter desired number:
7
```

Exit nested design

```
Enter number of desired funtion:
4
```

Return to BSDM

```
SELECT ANY KEY
SELECT ANY KEY
Select option desired :
1
```

Select Transform key

Algebraic transformation

```
Transformation number = ?
1
Variable number corresponding to X = ?
1
Parameter a = ?
1
Parameter b = ?
1
Parameter c = ?
1
Store transformed data in Variable # ( (= 2 )
?
2
Variable name ((= 10 characters) = ?
LN(Y+1)
Is above information correct?
YES
Press 'CONTINUE' when ready.
```

```
The following transformation was performed: a*(X^b)+c
  where a =  1
        b =  1

        c =  1
        X is Variable # 1
        Transformed data is stored in Variable # 2 (LN(Y+1)).
```

```
Select option desired :
1
Transformation number = ?
3
Variable number corresponding to X = ?
2
Parameter a = ?
1
Parameter b = ?
1
Parameter c = ?
0
Store transformed data in Variable # ( (= 3 )
?
2
Is above information correct?
YES
Press 'CONTINUE' when ready.
```

Another algebraic transformation

```
The following transformation was performed: a*ln(bX)+c
  where a =  1
        b =  1

        c =  0
        X is Variable # 2
        Transformed data is stored in Variable # 2 (LN(Y+1)).


Select option desired :
0
PROGRAM NOW UPDATING SCRATCH DATA FILE
SELECT ANY KEY
```

Exit transformation routine

Select LIST key

```
Option number = ?
1
Enter method for listing data:
3
```

```
                   THANKSGIVING GREETING CARD EVALUATION

Data type is: Raw data


       Variable # 1   Variable # 2
       (DESIGN     )   (LN(Y+1)    )

OBS#
   1       1.21000          .79299
   2       1.49000          .91228
   3       1.76000         1.01523
   4       1.52000          .92426
   5        .65000          .50078
   6       1.96000         1.08519
   7       1.21000          .79299
   8       1.57000          .94391
   9       1.72000         1.00063
  10       2.09000         1.12817
  11       2.21000         1.16627
  12       2.36000         1.21194
  13       2.83000         1.34286
  14       3.99000         1.60744
```

15	2.01000	1.10194
16	2.62000	1.28647
17	1.72000	1.00063
18	1.44000	.89200
19	1.84000	1.04380
20	.91000	.64710
21	1.30000	.83291
22	7.61000	2.15292
23	2.01000	1.10194
24	3.27000	1.45161
25	.29000	.25464
26	.92000	.65233
27	.37000	.31481
28	.72000	.54232
29	.43000	.35767
30	3.99000	1.60744
31	2.35000	1.20896
32	4.71000	1.74222
33	1.44000	.89200
34	2.09000	1.12817
35	1.84000	1.04380
36	2.36000	1.21194
37	1.96000	1.08519
38	3.26000	1.44927
39	2.01000	1.10194
40	1.70000	.99325
41	4.43000	1.69194
42	3.66000	1.53902
43	.51000	.41211
44	1.78000	1.02245
45	2.13000	1.14103
46	5.58000	1.88403
47	1.41000	.87963
48	2.75000	1.32176
49	2.60000	1.28093
50	2.21000	1.16627
51	1.44000	.89200
52	1.20000	.78846
53	1.21000	.79299
54	3.03000	1.39377
55	2.79000	1.33237
56	1.18000	.77932
57	1.67000	.98208
58	1.16000	.77011
59	1.73000	1.00430
60	1.92000	1.07158
61	4.84000	1.76473
62	2.88000	1.35584
63	4.10000	1.62924
64	1.48000	.90826
65	3.67000	1.54116
66	.78000	.57661
67	1.46000	.90016
68	1.65000	.97456
69	3.23000	1.44220
70	1.92000	1.07158
71	4.51000	1.70656
72	1.48000	.90826
73	1.33000	.84587
74	.39000	.32930
75	1.33000	.84587
76	1.37000	.86289
77	2.02000	1.10526
78	1.68000	.98582
79	4.51000	1.70656
80	2.34000	1.20597
81	3.33000	1.46557

82	1.16000	.77011
83	1.86000	1.05082
84	1.92000	1.07158
85	3.23000	1.44220
86	2.64000	1.29198
87	3.96000	1.60141
88	2.22000	1.16938
89	4.67000	1.73519
90	1.90000	1.06471
91	2.61000	1.28371
92	3.27000	1.45161
93	2.26000	1.18173
94	2.36000	1.21194
95	2.30000	1.19392
96	1.55000	.93609

```
Option number = ?
0                                                   Exit list procedure
SELECT ANY KEY
                                                    Return to AOV2

Enter number of desired funtion:
2                                                   Select nested design
Number of factors in design ? (2, 3, or 4)
3
Number of levels of factor A
?
2
Number of levels of factor B
?
6
Number of levels of factor C
?
8
Number of samples ?
1
Is the above information correct ?
YES
Which design (by number) is to be used ?
3
Which factor is P: A,B,C
?
A
Which factor is Q: B,C
?
C
Do YOU want to assign names to the factors ?
YES
Enter the name for factor A (<11 characters)
?
DISPLAY
Enter the name for factor B (<11 characters)
?
CARD STYLE
Enter the name for factor C (<11 characters)
?
STORES
Which variable number contains the response ?
2
No. of decimal places to print calculated values.
4
```

```
*******************************************************************************
                          NESTED ANALYSIS OF VARIANCE
*******************************************************************************
                     THANKSGIVING GREETING CARD EVALUATION

DESIGN
------
 Number of factors =  3
    No. of levels of factor A =  2
    No. of levels of factor B =  6
    No. of levels of factor C =  8
    No. of minor replications (samples) =  1

Response variable(s) are :
Variable no.  2      LN(Y+1)

MEANS
-----
* Overall mean =          1.1068


* Main Effect Means :

Factor A - DISPLAY   Levels ( 1 - 2 ) :
         1.0711          1.1426
Factor B - CARD STYLE   Levels ( 1 - 6 ) :
          .9621          1.2082          1.1403           .9105          1.1730
         1.2469
Factor C - STORES   Levels ( 1 - 8 ) :
         1.1236           .9108           .9144           .9817          1.0825
         1.4248          1.2798          1.1372


* Two Way Interaction Means :

Factor A - DISPLAY  down and  Factor B - CARD STYLE  across
                    1               2               3               4
                    5               6
    1                .8710          1.2307          1.1404           .8350
                    1.1132          1.2365
    2               1.0533          1.1858          1.1401           .9859
                    1.2329          1.2574

Factor A - DISPLAY  down and  Factor C - STORES  across
                    1               2               3               4
                    5               6               7               8
    1                .9388          1.0420           .8327           .9267
                     .8767          1.6310          1.0312          1.2899
    2               1.3085           .7795           .9961          1.0368
                    1.2882          1.2185          1.5283           .9845

Factor B - CARD STYLE  down and  Factor C - STORES  across
                    1               2               3               4
                    5               6               7               8
    1               1.0370          1.0393           .9536           .8564
                     .6469          1.2395          1.0627           .8616
    2                .9914           .9491          1.0853          1.1418
                    1.5538          1.4816          1.3656          1.0974
    3               1.2709           .7343           .9720           .8108
                    1.1376          1.6123          1.4043          1.1799
    4                .5503           .4908           .5803           .7026
                     .7315          1.2966          1.4578          1.4741
```

```
  5          1.1788      .9491     1.0473     1.1418
             1.2637     1.3706     1.3517     1.0813
  6          1.7136     1.3019      .8479     1.2370
             1.1614     1.5480     1.0368     1.1289

Should the 3-way means be printed ?
YES


* Three Way Interaction Means :
Factor A - DISPLAY,  Level 1
Factor B - CARD STYLE  down and  Factor C - STORES  across
             1          2          3          4
             5          6          7          8
  1           .7930      .9123     1.0152      .9243
              .5008     1.0852      .7930      .9439
  2          1.0006     1.1282     1.1663     1.2119
             1.3429     1.6074     1.1019     1.2865
  3          1.0006      .8920     1.0438      .6471
              .8329     2.1529     1.1019     1.4516
  4           .2546      .6523      .3148      .5423
              .3577     1.6074     1.2090     1.7422
  5           .8920     1.1282     1.0438     1.2119
             1.0852     1.4493     1.1019      .9933
  6          1.6919     1.5390      .4121     1.0225
             1.1410     1.8840      .8796     1.3218

Factor A - DISPLAY,  Level 2
Factor B - CARD STYLE  down and  Factor C - STORES  across
             1          2          3          4
             5          6          7          8
  1          1.2809     1.1663      .8920      .7885
              .7930     1.3938     1.3324      .7793
  2           .9821      .7701     1.0043     1.0716
             1.7647     1.3558     1.6292      .9083
  3          1.5412      .5766      .9002      .9746
             1.4422     1.0716     1.7066      .9083
  4           .8459      .3293      .8459      .8629
             1.1053      .9858     1.7066     1.2060
  5          1.4656      .7701     1.0508     1.0716
             1.4422     1.2920     1.6014     1.1694
  6          1.7352     1.0647     1.2837     1.4516
             1.1817     1.2119     1.1939      .9361
```

Note: Below AOV table does not show F ratios because the appropriate error mean square depends on the design.

```
ANOVA TABLE
-----------
                     Nested Analysis of Variance

Source (Name)     df     Sums of Squares        Mean Square
-----------------------------------------------------------
Total             95          12.5531              .1321
A    DISPLAY       1            .1225              .1225
C(A)              14           5.3373              .3812
B    CARD STYLE    5           1.5185              .3037
AB                 5            .1687              .0337
CB(A)             70           5.4062              .0772
```

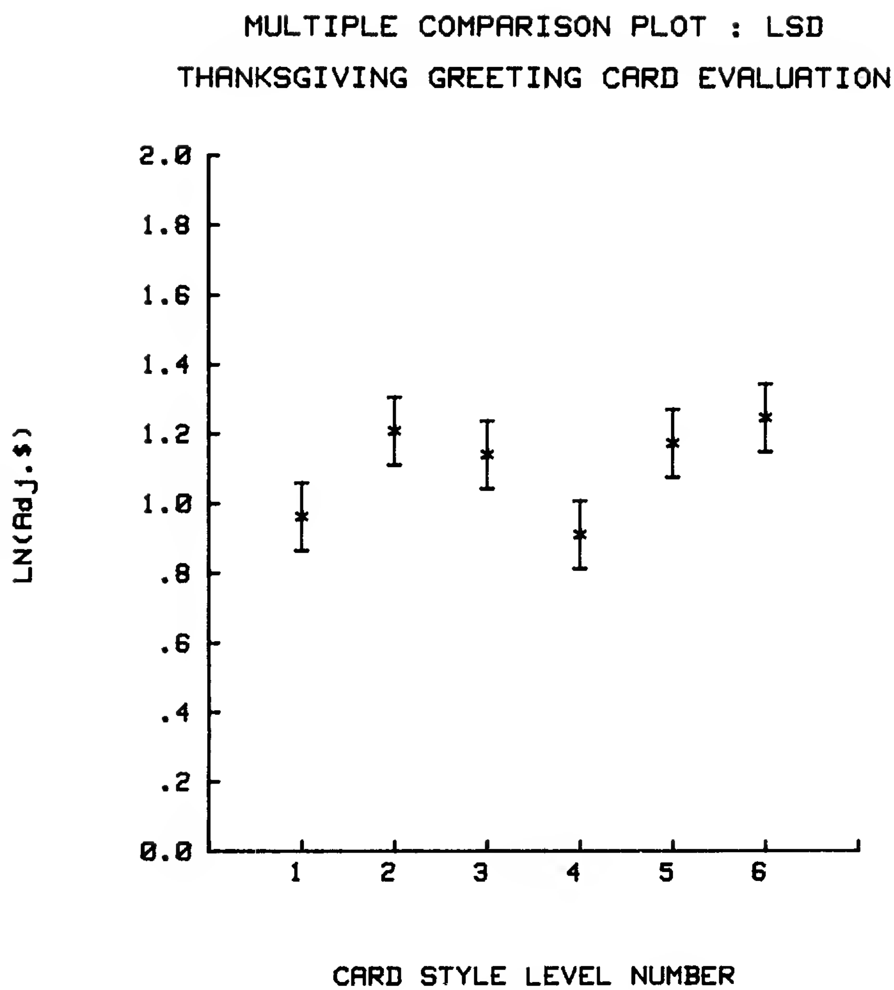
F = 3.93

This table shows the differences among card styles are even more significant.

```
Specify a new variable for this design ?
NO

Enter desired number:
3
Is the design displayed on the CRT the latest one?
YES
```

Multiple comparisons

```
*******************************************************************************
                         Multiple Comparisons
*******************************************************************************
Enter 1 or 2 to specify type of means
1                                                        Least significant difference
Which Factor/Main Effect(A,B, or C)should be used?
B
Error Mean Square, associated Degrees of Freedom
.07723,70
Which procedure would you like to use ?
1
What level of Alpha are you going to use ?
.05
Enter table value from Student's t with d.f.= 70
?
1.99
Is a plot of LSD desired ?
YES
Plot on CRT ?
NO
Plotter indentifier string (press CONT if 'HPGL')?

Enter select code, bus # (defaults are 7,5)?

Which PEN color should be used?
1
Enter name for labelling Y axis (( 11 characters)
LN(Adj.$)
Beep will sound when plot done, then press CONT

To interrupt plotting press 'STOP' key.
```




```
Least Significant Difference

  Error mean square = .07723
  Degrees of freedom = 70
  Harmonic average sample size =   16.0000
  Alpha level = .05
  Table value from Student's t = 1.99
  LSD value =    .1955
```

Multiple Comparisons on Factor CARD STYLE

Level	Mean	Sample Size	Separation
4	.9105	16	a
1	.9621	16	ab
3	1.1403	16	bc
5	1.1730	16	c
2	1.2082	16	c
6	1.2469	16	c

Note: Where the 'levels' do not contain the same letters the factor levels are significantly different using the LSD procedure.

```
Another Separation Procedure on Factor 2
?
NO
Another Factor to be used ?
NO
Multiple Comparison Procedures on Two-Way Means ?
NO

Enter number of desired funtion:
9                                                    Return to BSDM
```

Split Plot Example

Example

Hicks (1973, ex. 13.1) describes a split-plot experiment in which four oven temperatures and three baking times were investigated with regard to the life, Y, of an electrical component. The oven temperatures and the replications (blocks) are in the whole plot while the baking times are in the subplots. Only one electrical component was subjected to the stress conditions within each block-baking time-temperature combination.

The data table is shown below:

	Baking Time (A)	Oven Temp. (B) 580	600	620	640
Replication 1	5	217	158	229	223
	10	233	138	186	227
	15	175	152	155	156
2	5	188	126	160	201
	10	201	130	170	181
	15	195	147	161	172
3	5	162	122	167	182
	10	110	185	181	201
	15	113	180	182	199

Since this is a balanced design with three replications, we need only use one variable for data entry. The data is entered across each row in the table above. Hence, three groups of replications are available with factor A as baking time and factor B as oven temperature.

Within the split-plot program, we answer that there are two factors and three major replications. The design is specified with factor B in the whole plot and factor A in the subplot. The F ratio shows only significant temperature effects (B). The HSD multiple comparison procedure suggests that oven temperature two is significantly lower in life time readings than are the other three temperatures.

This conclusion is supported, as should be expected, by the more 'liberal' LSD procedure shown on the next multiple comparison output.

If one runs this data set through the Factorial Analysis in order to separate the replication interaction terms as suggested by Hicks, one finds a highly questionable interaction between replications and baking time. To do this, you specify factor A as replication, factor B as baking time, and factor C as oven temperature in the FACTORIAL program.

Note that in Hicks the printed AOV table shows the mean square for AB (replication by baking time) is 1755.32 which is substantially larger than any of the other replication interactions.

After looking at the data set, we believe that Hicks may have rearranged the original data, since you would ordinarily not expect the replication interaction terms to differ by that much in a split plot. See if you agree.

```
******************************************************************************
*                              DATA MANIPULATION                             *
******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                         Raw data
Mode number = ?
2                                                         On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
HICKS:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


               HICKS SPLIT PLOT ON COMPONENT LIFE TIME


Data file name: HICKS:INTERNAL

Data type is:   Raw data

Number of observations:    36
Number of variables:        1


Variable names:
   1. LIFETIME

Subfiles:  NONE


SELECT ANY KEY
                                                 Select special function key labeled-LIST
Option number = ?
1

                 HICKS SPLIT PLOT ON COMPONENT LIFE TIME

Data type is: Raw data


                              VARIABLE # 1  (LIFETIME)
```

I	OBS(I)	OBS(I+1)	OBS(I+2)	OBS(I+3)	OBS(I+4)
1	217.00000	158.00000	229.00000	223.00000	233.00000
6	138.00000	186.00000	227.00000	175.00000	152.00000
11	155.00000	156.00000	188.00000	126.00000	160.00000
16	201.00000	201.00000	130.00000	170.00000	181.00000
21	195.00000	147.00000	161.00000	172.00000	162.00000
26	122.00000	167.00000	182.00000	170.00000	185.00000

```
  31      181.00000       201.00000       213.00000       180.00000       182.00000
  36      199.00000
Option number = ?
0
SELECT ANY KEY

Enter number of desired funtion:
3
Number of factors in design ? (2 or 3)
2
Number of levels of factor A
?
3
Number of levels of factor B
?
4
Number of blocks in this design ?
3
No. obs per trt combination in each block(sample)?
1
Do YOU want to assign names to the factors ?
YES
Enter the name for factor A (<11 characters)
?

BAKINGTIME
Enter the name for factor B (<11 characters)
?
OVEN TEMP.
Which factor(s) are in the whole plots ?
B
Which factor(s) are in the split plots ?
A
Is the above information correct ?
YES
No. of decimal places to print calculated values.
4
```

Select special function key labeled-ADV STAT
Remove BSDM media
Insert AOV2 media

Split plot designs

```
********************************************************************************
                      SPLIT PLOT ANALYSIS OF VARIANCE
********************************************************************************
                   HICKS SPLIT PLOT ON COMPONENT LIFE TIME

DESIGN
------
   Number of factors =  2
    No. of levels of factor A =  3
    No. of levels of factor B =  4
    No. of major replications (blocks) =  3
    No. of minor replications (samples) =  1

Subfiles will be ignored
Whole plot factor(s) are :
     Factor B
Split-plot factor(s) are :
     Factor A
Response variable(s) are :
Variable no.  1      LIFETIME

MEANS
-----
* Overall mean =        178.4722
```

```
* Block and Main Effect Means :

Factor Blocks -      Levels ( 1 - 3 ) :
       187.4167       169.3333       178.6667
Factor A - BAKINGTIME   Levels ( 1 - 3 ) :
       177.9167       183.5833       173.9167
Factor B - OVEN TEMP.   Levels ( 1 - 4 ) :
       194.8889       148.6667       176.7778       193.5556


* Two Way Interaction Means :

Factor A - BAKINGTIME  down and  Factor B - OVEN TEMP.  across
                  1             2             3             4
   1          189.0000      135.3333      185.3333      202.0000
   2          201.3333      151.0000      179.0000      203.0000
   3          194.3333      159.6667      166.0000      175.6667



ANOVA TABLE
-----------
                 Split Plot Analysis of Variance

Source (Name)    df    Sums of Squares       Mean Square     F Ratio  F-Prob
--------------------------------------------------------------------------------
Total            35        29330.9722          838.0278

Blocks            2         1962.7222          981.3611      3.319    .1070
B  OVEN TEMP.     3        12494.3056         4164.7685     14.086    .0040
Error (a)         6         1773.9444          295.6574


A  BAKINGTIME     2          566.2222          283.1111       .456    .6418
BA                6         2600.4444          433.4074       .698    .6551
Error (b)        16         9933.3333          620.8333

    NOTE: F tests assume that all factors are fixed
```

Only factor B has a significant difference among effects.

```
Enter desired number:
1
```

Orthogonal polynomial comparisons

```
Is the design displayed on the CRT the latest one?
YES




********************************************************************************

                  Orthogonal Polynomial Comparisons

********************************************************************************
Orthogonal polynomial comparisons on FACTOR 1
?
YES
Enter the max degree of orthogonal poly
2
Value associated with level # 1 of FACTOR 1
?
5
```

```
Value associated with level # 2 of FACTOR 1
?
10
Value associated with level # 3 of FACTOR 1
?
15
Is the above information correct ?
YES
Enter Error mean square, degrees of freedom
620.83,16
```

From AOV table

Orthogonal Polynomial Decomposition on BAKINGTIME

Degree	SS	F-Ratio	F-Prob
1	96.0000	.1546	.69934
2	470.2222	.7574	.39701

```
Level of Treatments :  5  10  15
Orthogonal poly comparisons on another FACTOR?
YES
Orthogonal polynomial comparisons on FACTOR 1
?
NO
Orthogonal polynomial comparisons on FACTOR 2
?
YES
Enter the max degree of orthogonal poly
3
Value associated with level # 1 of FACTOR 2
?
580
Value associated with level # 2 of FACTOR 2
?
600
Value associated with level # 3 of FACTOR 2
?
620
Value associated with level # 4 of FACTOR 2
?
640
Is the above information correct ?
YES
Enter Error mean square, degrees of freedom
295.66,6
```

From AOV table

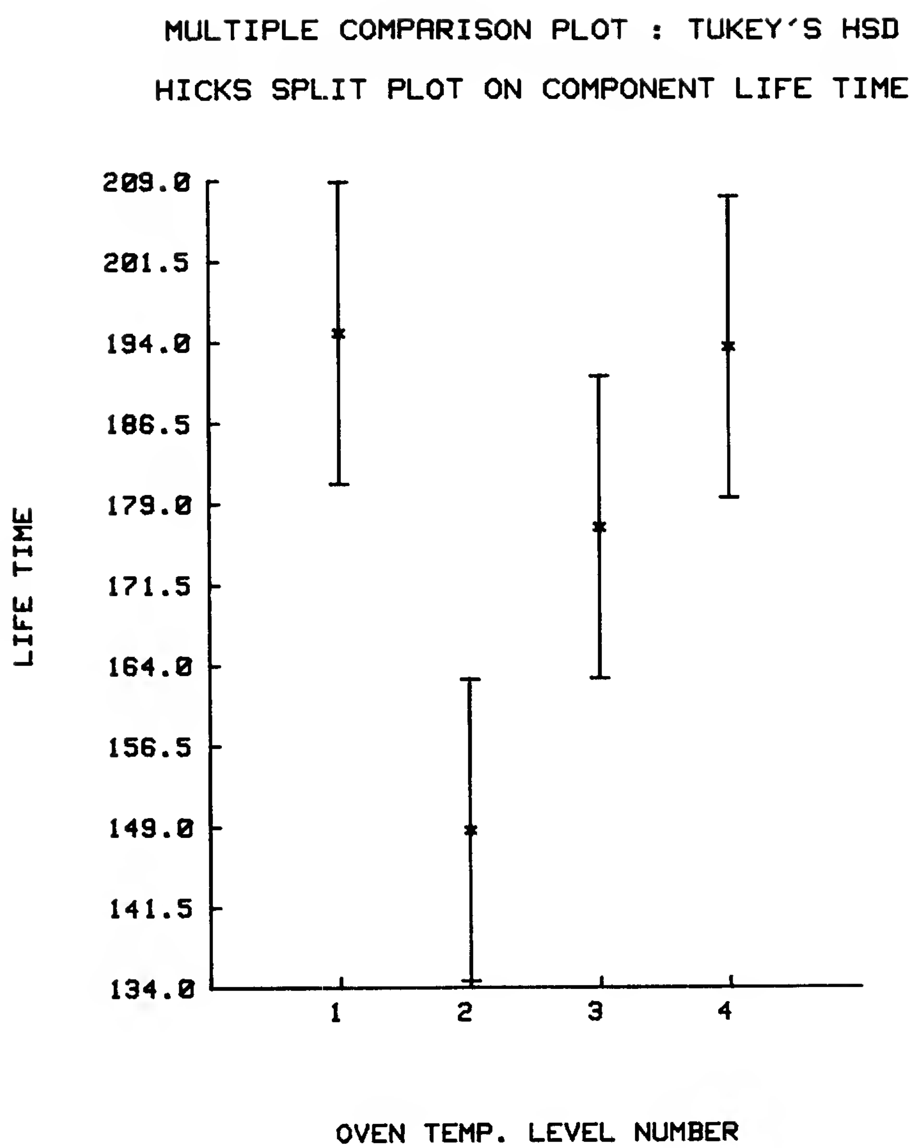
Orthogonal Polynomial Decomposition on OVEN TEMP.

Degree	SS	F-Ratio	F-Prob
1	261.6056	.8848	.38320
2	8930.2500	30.2045	.00152
3	3302.4500	11.1698	.01557

```
Level of Treatments :  580  600  620  640
Orthogonal poly comparisons on another FACTOR?
NO

Enter number of desired funtion:
6
Is the design displayed on the CRT the latest one?
YES
```

Multiple comparisons

```
******************************************************************************

                         Multiple Comparisons

******************************************************************************

Enter 1 or 2 to specify type of means
1
Which Factor/Main Effect(A or B)should be used ?
B
Error Mean Square, associated Degrees of Freedom
295.66,6
Which procedure would you like to use ?
2                                                        Tukey's HSD
What level of Alpha are you going to use ?
.05
for 4 means, d.f.= 6
?
4.9
Is a plot of HSD desired ?
YES
Plot on CRT ?
NO
Plotter indentifier string (press CONT if 'HPGL')?

Enter select code, bus # (defaults are 7,5)?

Which PEN color should be used?
1
Enter name for labelling Y axis (< 11 characters)
LIFE TIME
Beep will sound when plot done, then press CONT

To interrupt plotting press 'STOP' key.
```



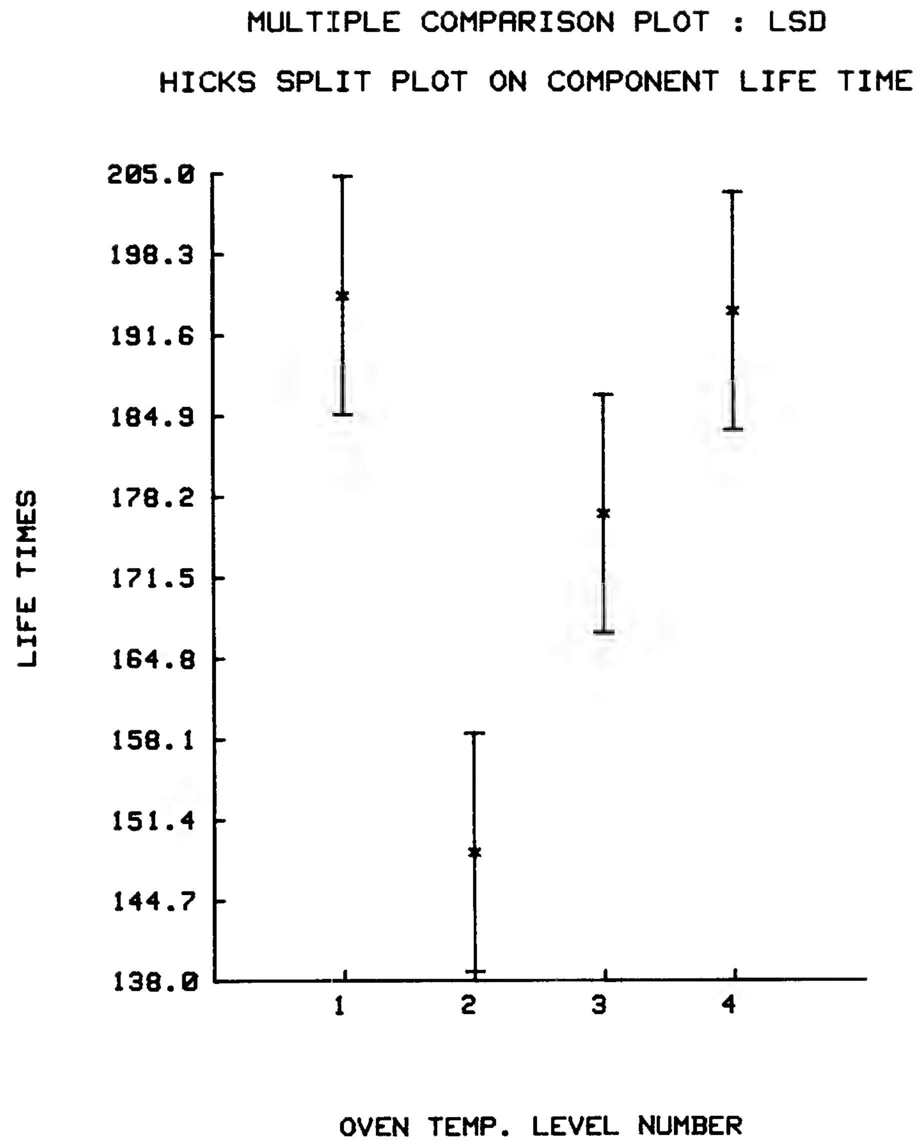
Tukey's HSD

Error mean square = 295.66
Degrees of freedom = 6
Harmonic average sample size = 9.0000
Alpha level = .05
Table value from Studentized range = 4.9
HSD value = 28.0848

Multiple Comparisons on Factor OVEN TEMP.

Level	Mean	Sample Size	Separation
2	148.6667	9	a
3	176.7778	9	b
4	193.5556	9	b
1	194.8889	9	b

```
Another Separation Procedure on Factor 2
?
NO
Another Factor to be used ?
NO
Multiple Comparison Procedures on Two-Way Means ?
NO

Enter number of desired funtion:
6                                                       Multiple comparisons
Is the design displayed on the CRT the latest one?
YES



****************************************************************************

                         Multiple Comparisons

****************************************************************************

Enter 1 or 2 to specify type of means
1
Which Factor/Main Effect(A or B)should be used ?
B
Error Mean Square, associated Degrees of Freedom
295.66,6
Which procedure would you like to use ?
1                                                       Least significant difference
What level of Alpha are you going to use ?
.05
Enter table value from Student's t with d.f.= 6
?
2.447
Is a plot of LSD desired ?
YES
Plot on CRT ?
NO
Plotter indentifier string (press CONT if 'HPGL')?

Enter select code, bus # (defaults are 7,5)?

Which PEN color should be used?
1
Enter name for labelling Y axis (< 11 characters)
LIFE TIMES
Beep will sound when plot done, then press CONT

To interrupt plotting press 'STOP' key.
```



Least Significant Difference

Error mean square = 295.66
Degrees of freedom = 6
Harmonic average sample size = 9.0000
Alpha level = .05
Table value from Student's t = 2.447
LSD value = 19.8346

Multiple Comparisons on Factor OVEN TEMP.

Level	Mean	Sample Size	Separation
2	148.6667	9	a
3	176.7778	9	b
4	193.5556	9	b
1	194.8889	9	b

```
Another Separation Procedure on Factor 2
?
NO
Another Factor to be used ?
NO
Multiple Comparison Procedures on Two-Way Means ?
NO

Enter number of desired funtion:
1                                                        Factorial design
Number of factors in design ? (2, 3, or 4)
3
Number of levels of factor A
?
3
Number of levels of factor B
?
3
Number of levels of factor C
?
4
Number of blocks in this design ?
1
No. obs per trt combination in each block(sample)?
1
Is the above information correct ?
YES
Do YOU want to assign names to the factors ?
YES
Enter the name for factor A (<11 characters)
?
REP
Enter the name for factor B (<11 characters)
?
BAKE TIME
Enter the name for factor C (<11 characters)
?
OVEN TEMP.
No. of decimals for printing calc. values(<=7).
4
*******************************************************************************
*                       FACTORIAL ANALYSIS OF VARIANCE                        *
*******************************************************************************
                    HICKS SPLIT PLOT ON COMPONENT LIFE TIME


DESIGN
------
   Number of factors =  3
    No. of levels of factor A =  3
    No. of levels of factor B =  3
    No. of levels of factor C =  4
    No. of major replications (blocks) =  1
    No. of minor replications (samples) =  1

Subfiles will be ignored
Response variable(s) are :
Variable no.  1      LIFETIME

MEANS
-----
* Overall mean =        178.4722
```

* Main Effect Means :

Factor A - REP   Levels (1 - 3) :

187.4167	169.3333	178.6667

Factor B - BAKE TIME   Levels (1 - 3) :

177.9167	183.5833	173.9167

Factor C - OVEN TEMP.   Levels (1 - 4) :

194.8889	148.6667	176.7778	193.5556

* Two Way Interaction Means :

Factor A - REP   down and   Factor B - BAKE TIME   across

	1	2	3
1	206.7500	196.0000	159.5000
2	168.7500	170.5000	168.7500
3	158.2500	184.2500	193.5000

Factor A - REP   down and   Factor C - OVEN TEMP.   across

	1	2	3	4
1	208.3333	149.3333	190.0000	202.0000
2	194.6667	134.3333	163.6667	184.6667
3	181.6667	162.3333	176.6667	194.0000

Factor B - BAKE TIME   down and   Factor C - OVEN TEMP.   across

	1	2	3	4
1	189.0000	135.3333	185.3333	202.0000
2	201.3333	151.0000	179.0000	203.0000
3	194.3333	159.6667	166.0000	175.6667

Should the 3-way means be printed ?
YES

* Three Way Interaction Means :

Factor A - REP,   Level 1
Factor B - BAKE TIME   down and   Factor C - OVEN TEMP.   across

	1	2	3	4
1	217.0000	158.0000	229.0000	223.0000
2	233.0000	138.0000	186.0000	227.0000
3	175.0000	152.0000	155.0000	156.0000

Factor A - REP,   Level 2
Factor B - BAKE TIME   down and   Factor C - OVEN TEMP.   across

	1	2	3	4
1	188.0000	126.0000	160.0000	201.0000
2	201.0000	130.0000	170.0000	181.0000
3	195.0000	147.0000	161.0000	172.0000

Factor A - REP,   Level 3
Factor B - BAKE TIME   down and   Factor C - OVEN TEMP.   across

	1	2	3	4
1	162.0000	122.0000	167.0000	182.0000
2	170.0000	185.0000	181.0000	201.0000
3	213.0000	180.0000	182.0000	199.0000

ANOVA TABLE

```
                    Factorial Analysis of Variance

Source (Name)     df    Sums of Squares         Mean Square
-----------------------------------------------------------
Total             35         29330.9722            838.0278
A    REP           2          1962.7222            981.3611
B    BAKE TIME     2           566.2222            283.1111
C    OVEN TEMP.    3         12494.3056           4164.7685
AB                 4          7021.2778           1755.3194
AC                 6          1773.9444            295.6574
BC                 6          2600.4444            433.4074
ABC               12          2912.0556            242.6713
```

We can see that the interaction between baking temperature and replication is significant.

```
Enter desired number:
7
```

Exit factorial design.

```
Enter number of desired funtion:
4
```

Return to BS DM.

One Way AOV

Example

Tissue Culture Growth was studied after exposure to five 'sugar' treatments; control, 2% fructose, 1% glucose and 1% fructose, and 2% sucrose. The response, Y, is length (in ocular units) of pea section grown in tissue culture with auxin present.

The data was entered using One-Way AOV mode 2 in which all treatments are stored in one variable. Each treatment has ten observations (samples). Hence, observations 1 to 10 are in the first treatment, observations 11 to 20 are in the second treatment, etc. The F ratio for treatments shows a very strong indication that the population treatment levels are significantly different. Both the LSD and Duncan Multiple Comparison procedure separate the treatments into three non-overlapping groups - treatments 4, 3, and 2: and treatment 5; and treatment 1 (control). Hence, if you add either glucose (2) or fructose (3) or both (4) you get shorter lengths that if you use just sucrose which is in turn shorter than the control treatment.

```
********************************************************************************
*                              DATA MANIPULATION                               *
********************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                      Raw data
Mode number = ?
2                                                      On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
TISSUE:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


                         TISSUE CULTURE GROWTH


Data file name: TISSUE:INTERNAL

Data type is:   Raw data

Number of observations:    50
Number of variables:        1


Variable names:
   1. GROWTH

Subfile name      beginning observation  number of observations
 1. CONTROL                           1                       10
 2. 2% GLUCOSE                       11                       10
 3. 2% FRUCT.                        21                       10
 4. 1%GLU+1FRU                       31                       10
 5. 2%SUCROSE                        41                       10
```

```
SELECT ANY KEY
```

Select special function key labeled-LIST

```
Option number = ?
1
```

List all data

```
                    TISSUE CULTURE GROWTH

Data type is: Raw data

                         VARIABLE # 1  (GROWTH)
```

I	OBS(I)	OBS(I+1)	OBS(I+2)	OBS(I+3)	OBS(I+4)
1	75.00000	67.00000	70.00000	75.00000	65.00000
6	71.00000	67.00000	67.00000	76.00000	68.00000
11	57.00000	58.00000	60.00000	59.00000	62.00000
16	60.00000	60.00000	57.00000	59.00000	61.00000
21	58.00000	61.00000	56.00000	58.00000	57.00000
26	56.00000	61.00000	60.00000	57.00000	58.00000
31	58.00000	59.00000	58.00000	61.00000	57.00000
36	56.00000	58.00000	57.00000	57.00000	59.00000
41	62.00000	66.00000	65.00000	63.00000	64.00000
46	62.00000	65.00000	65.00000	62.00000	67.00000

```
Option number = ?
0
```

Exit list procedure

```
SELECT ANY KEY
```

Select ADV STAT
Remove BSDM media
Insert AOV1 media
Select one way classification

```
Enter number of desired funtion:
1
How many treatments in this analysis ?
5
Enter name for treatment/factor (<11 characters)
TISSUE
Do YOU want to assign names to the treatments ?
YES
Enter the name for treatment 1 (<11 characters)
?
CONTROL
Enter the name for treatment 2 (<11 characters)
?
2% GLUCOSE
Enter the name for treatment 3 (<11 characters)
?
2% FRUCT.
Enter the name for treatment 4 (<11 characters)
?
1%GLU+FRU
Enter the name for treatment 5 (<11 characters)
?
2%SUCROSE
Are the names displayed on the CRT correct ?
YES
Treatment definition mode = ?
2
Enter the number of observations in treatment 1
?
10
Enter the number of observations in treatment 2
?
10
Enter the number of observations in treatment 3
?
10
```

```
Enter the number of observations in treatment 4
?
10
Enter the number of observations in treatment 5
?
10
Subfile # (enter 0 to ignore subfile) = ?
0
Is the design description on the CRT correct ?
YES

*****************************************************************************

                    ONE-WAY ANALYSIS OF VARIANCE:
                        TISSUE CULTURE GROWTH

*****************************************************************************

# of decimals for printing calculated values(<=7)?
4
DESIGN

    # of treatments =  5
    # of observations in treatment 1  =  10
    # of observations in treatment 2  =  10
    # of observations in treatment 3  =  10
    # of observations in treatment 4  =  10
    # of observations in treatment 5  =  10
  Response = GROWTH


SUMMARY STATISTICS

                        Treatment Statistics

  Treatment name          Total            Mean        Stan.Dev    N
   CONTROL               701.0000         70.1000        3.9847   10
   2% GLUCOSE            593.0000         59.3000        1.6364   10
   2% FRUCT.             582.0000         58.2000        1.8738   10
   1%GLU+FRU             580.0000         58.0000        1.4142   10
   2%SUCROSE             641.0000         64.1000        1.7920   10

   Overall              3097.0000         61.9400        5.1958   50

*****************************************************************************

ANOVA TABLE

                  One-Way Analysis of Variance Table

  Source       Df         SS             MS         F-Ratio      F-Prob
  Total        49     1322.8200
  TISSUE        4     1077.3200       269.3300      49.3680     0.00000
  Error        45      245.5000         5.4556

*****************************************************************************
```

We can see that the effects of population treatment levels are significantly different.

```
Bartlett's test of homogeneity of variance :
    Chi-square value =   13.939 with degrees of freedom =     4
Do you wish to specify another subfile ?
NO
```

$X^2(4,.05) = 9.488, X^2(4,.01) = 13.277$

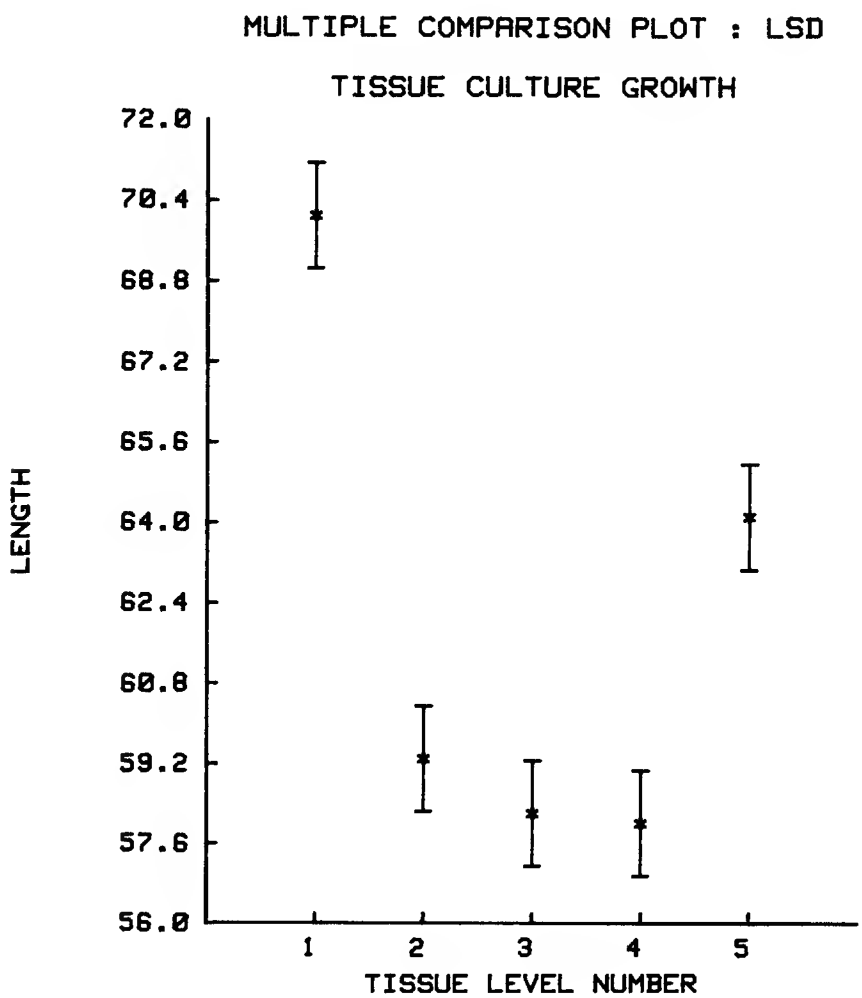
Both are smaller than the calculated X^2 value of 13.9386, so we know that the variances are not homogeneous.

```
Enter desired number:
3
```

Multiple comparisons

```
Is the design displayed on the CRT the latest one?
YES
```

```
*********************************************************************************
                              MULTIPLE COMPARISONS
*********************************************************************************
Which procedure would you like to use ?
1                                                             Least significant difference
What level of Alpha are you going to use ?
.05
Enter table value form Student's t with d.f= 45
?
2.014
Is a plot of LSD desired ?
YES
Plot on CRT ?
NO
Plotter indentifier string (press CONT if 'HPGL')?
Plotter select code, bus # (defaults are 7,5)?

Beep will sound when plot done, then press CONT.
Which PEN color should be used?
1
Enter name for labelling Y axis(<11 characters)
LENGTH
To interrupt plotting, press 'STOP' key.
```



```
Least Significant Difference

  Error mean square =    5.4556
  Degrees of freedom = 45
  Harmonic average sample size =   10.0000
  Alpha level = .05
  Table value from Student's t = 2.014
  LSD value =   2.1037
```

```
                    Multiple Comparisons on TISSUE

    Level          Mean           Sample Size          Separation
      4           58.0000             10                    a
      3           58.2000             10                    a
      2           59.3000             10                    a
      5           64.1000             10                    b
      1           70.1000             10                    c
```

This separates the treatment into three non-overlapping groups, treatments 4, 3, and 2 in one group, 5 in another, and 1 in the last.

```
Another Separation Procedure on TISSUE
?
YES
Which procedure would you like to use ?
3
What level of Alpha are you going to use ?
 05
```

Select Duncan's Test

```
Duncan's Test

  Error mean square =    5.4556
  Degrees of freedom = 45
  Harmonic average sample size =   10.0000
  Alpha level = .05

   Means Separated            Table Value          Required Difference
for  5 means and d.f.= 45
?
3.16
             5                  3.1600                  2.3340
for  4 means and d.f.= 45
?
3.095
             4                  3.0950                  2.2860
for  3 means and d.f.= 45
?
3.005
             3                  3.0050                  2.2195
for  2 means and d.f.= 45
?
2.85
             2                  2.8500                  2.1051
```

```
                    Multiple Comparisons on TISSUE

    Level          Mean           Sample Size          Separation
      4           58.0000             10                    a
      3           58.2000             10                    a
      2           59.3000             10                    a
      5           64.1000             10                    b
      1           70.1000             10                    c
```

Same conclusion as above

```
Another Separation Procedure on TISSUE
?
NO


NOTE: HARMONIC AVER SAMPLE SIZE OF  10  USED
      IN CALCULATING THE MULTIPLE COMPARISONS.
Enter number of desired funtion:
9                                                Return to BSDM
```

Two Way (Unbalanced)

Example

The following data from Bancroft (1968, Ex. 1.3) is a two-way classification with factor A representing five different batches of silver and factor B representing two batches of iodine which are used to make silver iodine. The response, Y, is the reacting weights (coded). Apparently several samples were lost because the design is unbalanced.

Iodine

		I_1	I_2
	S_1	22	−1
		25	40
			18
	S_2	41	23
		41	13
Silver	S_3	29	
		20	
		37	
	S_4	49	61
		50	
	S_5	55	

The data is entered using two variables. Variable one is used to identify the rows and columns and variable two contains the response, Y. Hence, a value in variable one of 0301 indicates that the observation in variable two is from the third level of silver (A) and the first level of Iodine (B). The Two-Way Unbalanced routine is used with the method of fitting constants selected as the desired procedures because of the presence of empty cells. This analysis indicates that the sampled batches of silver do not support the hypothesis of equality for the population means.

The multiple comparison procedure by Student, Newman & Keuls (SNK) shows no separation between the five samples of silver. This probably can be explained by both the conservative nature of the SNK procedure and the fact that the AOV procedure uses an adjusted mean square for silver.

```
*******************************************************************************
*                              DATA MANIPULATION                              *
*******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                        Raw data
Mode number = ?
2                                                        On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
SLVRIODN:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES


               CODED REACTIN WEIGHTS OF SLIVER IODINE


Data file name: SLVRIODN:INTERNAL

Data type is:   Raw data

Number of observations:    16
Number of variables:        2


Variable names:
   1. ROW;COLUMN
   2. RWEIGHT

Subfiles:  NONE


SELECT ANY KEY
                                                         Select special function key labeled-LIST
Option number = ?
1
Enter method for listing data:
3


                 CODED REACTIN WEIGHTS OF SLIVER IODINE

Data type is: Raw data


       Variable # 1   Variable # 2
       (ROW;COLUMN)   (RWEIGHT    )

OBS#
   1       101.00000       22.00000
   2       101.00000       25.00000
   3       201.00000       41.00000
   4       201.00000       41.00000
   5       301.00000       29.00000
   6       301.00000       20.00000
   7       301.00000       37.00000
   8       401.00000       49.00000
   9       401.00000       50.00000
```

```
   10       501.00000        55.00000
   11       102.00000        -1.00000
   12       102.00000        40.00000
   13       102.00000        18.00000
   14       202.00000        23.00000
   15       202.00000        13.00000
   16       402.00000        61.00000
```

```
Option number = ?
0
SELECT ANY KEY

Enter number of desired funtion:
2
Data storage type =
2
Variable number for packed identification =
1
Enter # of rows, # of columns (separate by comma)
5,2
Do YOU wish to label the row and column factors ?
YES
Enter name of row factor (<11 characters)
SILVER
Enter name of column factor (<11 characters)
IODINE
Enter the variable number for response
2
Is the above information correct ?
YES
```

Exit list routine

Select special function key labeled-ADV STAT
Remove BSDM media
Insert AOV1 media

Two-way unbalanced design

```
*********************************************************************************

                      TWO-WAY UNBALANCED ANALYSIS OF VARIANCE:
                        CODED REACTIN WEIGHTS OF SLIVER IODINE

*********************************************************************************

# of decimal places for calculated values (<=7)?
4
DESIGN

   # of rows =  5
   # of columns =  2
   Response = RWEIGHT

SUMMARY STATISTICS
```

Subclass Statistics

Row	Column	Total	Mean	Stan.Dev	N
1	1	47.0000	23.5000	2.1213	2
1	2	57.0000	19.0000	20.5183	3
2	1	82.0000	41.0000	0.0000	2
2	2	36.0000	18.0000	7.0711	2
3	1	86.0000	28.6667	8.5049	3
4	1	99.0000	49.5000	.7071	2
4	2	61.0000	61.0000	0.0000	1
5	1	55.0000	55.0000	0.0000	1

****

Row Statistics

Row	Total	Mean	N
1	104.0000	20.8000	5
2	118.0000	29.5000	4
3	86.0000	28.6667	3
4	160.0000	53.3333	3
5	55.0000	55.0000	1

****

Column Statistics

Col	Total	Mean	N
1	369.0000	36.9000	10
2	154.0000	25.6667	6

****

ANOVA TABLE

Preliminary AOV (Test two way model)

Source	Df	SS	MS	F-Ratio	F-Prob
Total	15	4255.4375			
Subclass	7	3213.7708	459.1101	3.5260	.04908
Error	8	1041.6667	130.2083		

Preliminary AOV (Test for Interaction)

Source	Df	SS	MS	F-Ratio	F-Prob
Total	15	4255.4375			
Main	5	2722.2592	544.4518	4.1814	.03641
Int	2	491.5116	245.7558	1.8874	.21308
Error	8	1041.6667	130.2083		

****

Analysis of Variance (Method of Fitting Constants)

Source	Df	SS	MS	F-Ratio	F-Prob
Total	15	4255.4375			
SILVER	4	2572.3042	643.0760		
IODINE (Adj)	1	149.9550	149.9550	1.1517	.31450
IODINE	1	473.2042	473.2042		
SILVER (Adj)	4	2249.0550	562.2638	4.3182	.03749
Int	2	491.5116	245.7558		
Error	8	1041.6667	130.2083		

****

Enter desired number:
3 Multiple comparisons
Is the design displayed on the CRT the latest one?
YES

****

```
                              MULTIPLE COMPARISONS
*******************************************************************************
Enter 1 or 2 to specify type of means
1
Which Factor/Main Effect(A or B)should be used ?
A
Which procedure would you like to use ?
4                                                       Student Newman-Kevls
What level of Alpha are you going to use ?
.05

Student-Newman-Keuls Test

  Error mean square =  130.2083
  Degrees of freedom = 8
  Harmonic average sample size =    2.3622
  Alpha level = .05

   Means Separated            Table Value           Required Difference
for  5 means and d.f.= 8
?
4.89
             5                  4.8900                 36.3053
for  4 means and d.f.= 8
?
4.53
             4                  4.5300                 33.6325
for  3 means and d.f.= 8
?
4.04
             3                  4.0400                 29.9945
for  2 means and d.f.= 8
?
3.26
             2                  3.2600                 24.2035


                    Multiple Comparisons on SILVER

    Level           Mean          Sample Size          Separation
      1            20.8000             5                    a
      3            28.6667             3                    a
      2            29.5000             4                    a
      4            53.3333             3                    a
      5            55.0000             1                    a


 Another Separation Procedure on SILVER
 ?
 NO
 Another Factor to be used ?
 NO
 Multiple Comparison Procedures on Two-Way Means ?
 NO

 NOTE: HARMONIC AVER SAMPLE SIZE OF  2.36220472441  USED
       IN CALCULATING THE MULTIPLE COMPARISONS.
 Enter number of desired funtion:                       Return to BSDM
 9
```

One Way Analysis of Covariance

Example

An experiment to evaluate the effects of various growth stiumulants (X-4 on tomato seedlings was performed in which:

X = Initial length of seedling (m.m.)

Y = Growth in length (m.m.) during experiment

Stimulant X-4		Stimulant BC		Stimulant F32		Stimulant OX	
X	Y	X	Y	X	Y	X	Y
29	22	15	30	16	12	5	23
20	22	9	32	31	8	25	31
14	20	1	26	26	13	16	28
21	24	6	25	35	25	10	26
6	12	19	37	12	7	24	33

The data was entered using the first mode of storage for the covariance program. That is, each X,Y pair was stored in two variables and each of the four treatments used different variable pairs. Hence, for the Stimulant X-4, the initial length, X, was stored in Variable 1 and the growth, Y, was stored in Variable 2; while for the stimulant OX, the X value was stored in Variable 7 and the Y in Variable 8. Each variable has five observations.

The first part of the output from the One-way Covariance routines shows the within treatement statistics including totals, means, standard deviations, sample sizes, correlation coefficients, and regression coefficients. Note that the correlation coefficient and regression coefficient are for all of the data points taken together without regard to treatment group. Hence, it should not be surprising that no overall relationship exists between the X and Y variables. The test for homogeneity of regression coefficients confirms that we can accept the hypothesis that all treatment regression coefficients are essentially the same. The test for significance of pooled regression confirms that the relationship between the X and Y pooled across all treatments is significant (level = .0003).

Whereas the F ratio for treatment differences on the X's is non-significant (level = .12117), the F ratio on the original Y's is significant at the .00037 level. The analysis of covariance adjustment to the original data does not change the significance of the treatment effect ($\propto$ = .00000), but rather makes the difference in the means even more pronounced. This is shown by studying the "Table of Means" and noting the adjustment made in the original Y means after the use of the covariate X.

The use of the Tukey HSD multiple comparison procedure shows that stimulants one and three differ from all other stimulants, while no significant difference can be shown between two and four.

```
*******************************************************************************
*                               DATA MANIPULATION                             *
*******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                          Raw data
Mode number = ?
2                                                          On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
TOMATO:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES
```

```
        EFFECTS OF GROWTH STIMULANTS ON TOMATO SEEDLING LENGTHS


Data file name: TOMATO:INTERNAL

Data type is:   Raw data

Number of observations:      5
Number of variables:         8


Variable names:
   1. X-4:I
   2. X-4:G
   3. BC:I
   4. BC:G
   5. F32:I
   6. F32:G
   7. OX:I
   8. OX:G

Subfiles:  NONE


SELECT ANY KEY
                                                           Select LIST key
Option number = ?
1
Enter method for listing data:
3
```

```
        EFFECTS OF GROWTH STIMULANTS ON TOMATO SEEDLING LENGTHS

Data type is: Raw data
```

OBS#	Variable # 1 (X-4:I)	Variable # 2 (X-4:G)	Variable # 3 (BC:I)	Variable # 4 (BC:G)	Variable # 5 (F32:I)
1	29.00000	22.00000	15.00000	30.00000	16.00000
2	20.00000	22.00000	9.00000	32.00000	31.00000
3	14.00000	20.00000	1.00000	26.00000	26.00000
4	21.00000	24.00000	6.00000	25.00000	35.00000
5	6.00000	12.00000	19.00000	37.00000	12.00000

```
       Variable # 6     Variable # 7     Variable # 8
       (F32:G      )    (OX:I       )    (OX:G       )

OBS#
   1       12.00000          5.00000         23.00000
   2        8.00000         25.00000         31.00000
   3       13.00000         16.00000         28.00000
   4       25.00000         10.00000         26.00000
   5        7.00000         24.00000         33.00000
```

```
Option number = ?
0                                                    Exit list procedure
SELECT ANY KEY                                       Select ADV STAT key
                                                     Remove BSDM media
                                                     Insert AOV1 media
Enter number of desired funtion:
3                                                    One way analysis of covariance
How many treatments in this analysis ?
4
Enter a name for treatment/factor(<11 characters)
TREATMENT
Do YOU want to assign names to the treatments ?
YES
Enter the name for trt. 1 (<=10 characters)
?
X-4
Enter the name for trt. 2 (<=10 characters)
?
BC
Enter the name for trt. 3 (<=10 characters)
?
F32
Enter the name for trt. 4 (<=10 characters)
?
OX
Are the names displayed on the CRT correct ?
YES
Treatment definition mode = ?
1
Enter the X var., Y var. for treatment 1
?
1,2
Enter the X var., Y var. for treatment 2
?
3,4
Enter the X var., Y var. for treatment 3
?
5,6
Enter the X var., Y var. for treatment 4
?
7,8
Is the design description on the CRT correct ?
YES

********************************************************************************
                         ONE-WAY ANALYSIS OF COVARIANCE
               EFFECTS OF GROWTH STIMULANTS ON TOMATO SEEDLING LENGTHS

********************************************************************************

# of decimal places for calculated values(<=7) ?
4
```

DESIGN

```
  # of treatments =  4
    # of observations in treatment 1  =  5
    # of observations in treatment 2  =  5
    # of observations in treatment 3  =  5
    # of observations in treatment 4  =  5
  Covariate X = X-4:I
  Response  Y = X-4:G
```

SUMMARY STATISTICS

Treatment Statistics

Treatment		Total	Mean	Stan.Dev	N
X-4	X	90.0000	18.0000	8.5732	5
	Y	100.0000	20.0000	4.6904	5
BC	X	50.0000	10.0000	7.1414	5
	Y	150.0000	30.0000	4.8477	5
F32	X	120.0000	24.0000	9.7724	5
	Y	65.0000	13.0000	7.1764	5
OX	X	80.0000	16.0000	8.6891	5
	Y	141.0000	28.2000	3.9623	5
Overall	X	340.0000	17.0000	9.4088	20
	Y	456.0000	22.8000	8.5076	20

Within Treatment Regressions

Treatment	Corr.Coef.	Regression Coef.
X-4	.8331	.4558
BC	.8449	.5735
F32	.6310	.4634
OX	.9730	.4437
Overall	-.0487	-.0440

****

ANOVA TABLE

One-Way Analysis of Variance Table(X-Variable)

Source	Df	SS	MS	F-Ratio	F-Prob
Total	19	1682.0000			
Treatment	3	500.0000	166.6667	2.2561	.12117
Error	16	1182.0000	73.8750		

One-Way Analysis of Variance Table(Y-Variable)

Source	Df	SS	MS	F-Ratio	F-Prob
Total	19	1375.2000			
Treatment	3	924.4000	308.1333	10.9364	.00037
Error	16	450.8000	28.1750		

****

We can see that the effects of X-variables have no significant difference, but the effects of Y-variables are significantly different.

```
Test of homogeneity of regression coefficients :
     F-value =       .0538 with    3 and   12 degrees of freedom
     P(F)      .05) =  .98277
```

We consider all treatment regression coefficients are the same.

```
Test of significance of pooled regression coefficient :
     F-value =     21.8324 with 1 and   15 degrees of freedom
     P(F)   21.83) =  .00030
```

We can see that the relationship between X and Y pooled across all treatments is significant.

```
Pooled Regression Coefficient  =  .475465313029
Pooled Correlation Coefficient =  .7699

*********************************************************************************


                    One Way Analysis of Covariance Table

  Source        Df            SS                MS          F-Ratio        F-Prob
  Total         18         1371.9444
  Treatment      3         1188.3559         396.1186       32.3647       0.00000
  Error         15          183.5885          12.2392

*********************************************************************************
```

We can see that the effects of treatments are significantly different.

```
                              Table of Y Means

     Treatment name  Unadjusted Y Mean   Adjusted Y Mean        Stand. Dev     N
        X-4                  20.0000           19.5245             1.5646     5
        BC                   30.0000           33.3283             1.5646     5
        F32                  13.0000            9.6717             1.5646     5
        OX                   28.2000           28.6755             1.5646     5

*********************************************************************************

Do you want to change response for this subfile?
NO

Enter desired number:
3
Is the design displayed on the CRT the latest one?
YES
```

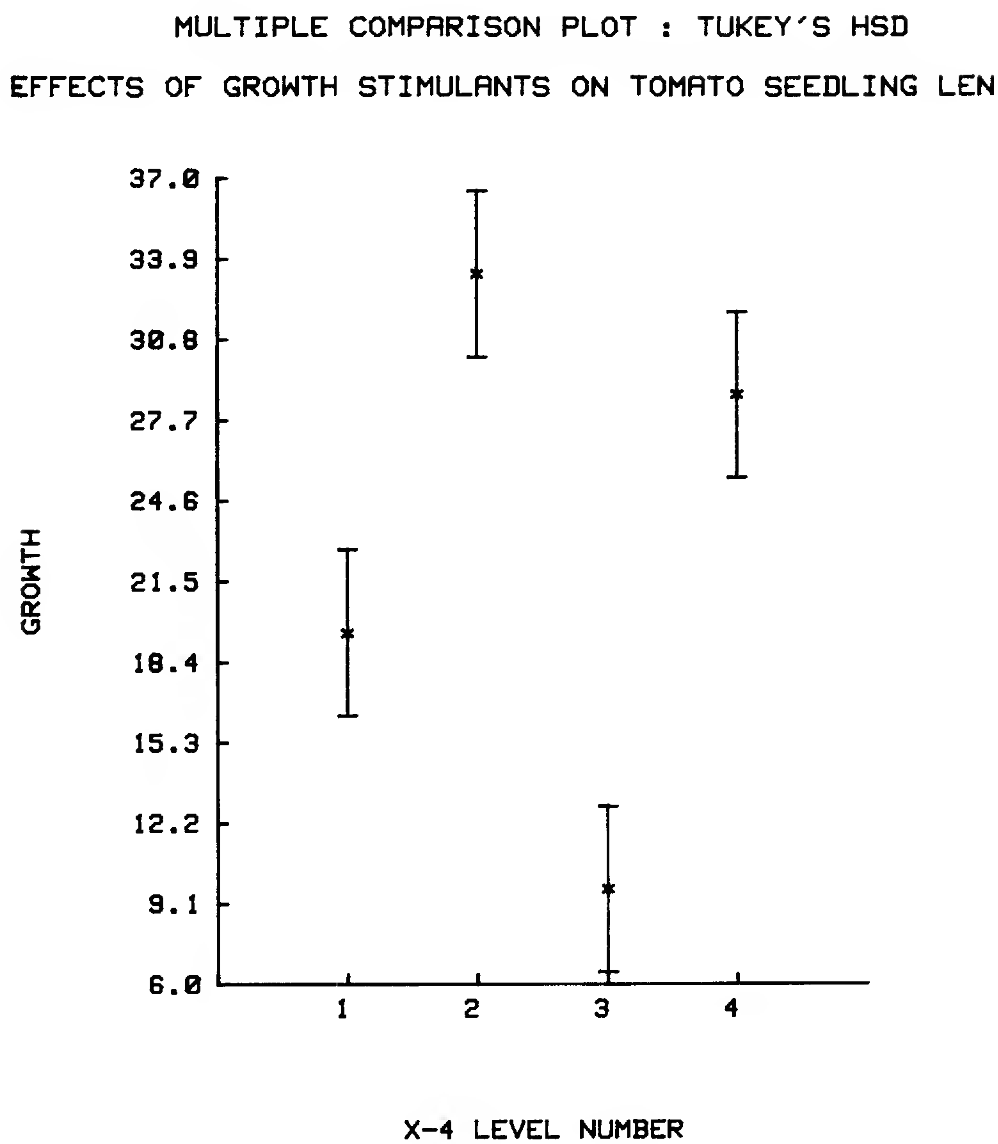
Multiple comparisons

```
*********************************************************************************

                           MULTIPLE COMPARISONS

*********************************************************************************

Which procedure would you like to use ?
2
```

Tukey's HSD

```
What level of Alpha are you going to use ?
.05
for 4 means and d.f.= 15
?
4.08
Is a plot of HSD desired ?
YES
Plot on CRT ?
NO
Plotter indentifier string (press CONT if 'HPGL')?

Plotter select code, bus # (defaults are 7,5)?
```

```
Beep will sound when plot done, then press CONT.
Which PEN color should be used?
1
Enter name for labelling Y axis(<11 characters)
GROWTH
To interrupt plotting, press 'STOP' key.
```




```
Tukey's HSD

  Error mean square =    12.2392
  Degrees of freedom = 15
  Harmonic average sample size =    5.0000
  Alpha level = .05
  Table value from Studentized range = 4.08
  HSD value =    6.3834
```

Level 3 differs from Level 1, which differs from Level 4 & 2

Multiple Comparisons on TREATMENT

Level	Mean	Sample Size	Separation
3	9.6717	5	a
1	19.5245	5	b
4	28.6755	5	c
2	33.3283	5	c

Another Separation Procedure on TREATMENT
?
NO

NOTE: HARMONIC AVER SAMPLE SIZE OF 5 USED
IN CALCULATING THE MULTIPLE COMPARISONS.
Enter number of desired funtion:
9 Return to BSDM

Notes

Principal Components and Factor Analysis

General Information

Description

The Principal Components and Factor Analysis Software accomplishes a variety of factor-analytic techniques. Input may be raw data, a correlation matrix, a covariance matrix, or a factor matrix. Factors are extracted from the correlation matrix. You may choose either the principal axes method or the maximum likelihood method to extract the initial factors. Orthogonal varimax or quartimax rotations and/or oblique oblimin rotations may be applied to the factor matrix. In the oblique rotation, you can control the degree of correlations among factors. Graphical presentation of the relationship between pairs of initial or rotated factors is also available.

The program computes the case scores and provides a plot for the case scores between each pair of factors if the raw data has been input. Case scores may be stored on a new file for further study.

For a brief discussion of the techniques and computing formulas used in these programs, see the Discussion Section.

Setting Up the Data

The first thing you need to do is to enter the data by using the Basic Statistics and Data Manipulation (BSDM) routines. The input may be the raw data, a correlation matrix, a covariance matrix, or a factor matrix. If a correlation matrix or a covariance matrix is to be entered, only the distinct elements will be requested, i.e., only the portion on and above the main diagonal. After the data has been loaded into memory, you are ready to use the Principal Components and Factor Analysis programs.

Special Considerations

Factor or Principal Component Scores

In the case where an observation has one or more missing values, the score for that observation will not be calculated and a blank line will be printed.

Storing the Correlation Matrix

In the case where it would be desirable to continue analysis at another time, you may store the correlation matrix. Note that the correlation matrix can later be input as data in BSDM.

Principal Components

Object of Program

A principal components analysis for a correlation matrix may be performed by selecting this option. Principal components will be printed. A table of eigenvalues is then printed. This includes the eigenvectors as well as the proportion and the cumulative proportion of the total variance accounted for by each component.

If raw data has been input, case scores on the components may be computed and stored. If a missing value is encountered in the calculation of component scores, the program will ignore that particular observation. Case scores are calculated for all observations in the data set even if the principal components were developed for only one subfile.

Special Considerations

Component Output Options

Four output options are available and are described on the CRT display. Each option allows you to inform the program how to determine how many components should be output. When using the minimum eigenvalue size option, many researchers choose a value of 1.00, while the maximum cumulative percent some researchers use is about 90 percent. The calculations, however, will be done for all principal components, i.e., one for each variable which has been included in the analysis. The number of components which result from your selected option will be used to determine the number printed later on in this routine.

Plots

For both the principal components plot and the component scores plot, you may select component numbers up to and including the number of variables you originally specified for the present analysis. Of course, if you originally had twenty variables, a plot of the 19th or 20th components may not be very useful.

Storing Principal Components Scores

The component scores are calculated and stored in the data matrix for all components which you specify. Component scores are generated for all observations in the data set across all subfiles. This feature may be useful for cross validation of the components between subfiles.

Factor Analysis

Object of Program

The extraction and rotation of the initial factors may be performed by selecting this option. Factors are extracted from a correlation matrix by the principal axes method or by the maximum likelihood method. If the principal axes method is used, three types of initial communality estimates may be used as diagonal elements of the correlation matrix; namely, squared multiple correlations, maximum absolute raw correlations or user-specified values.

For the principal axes method, you determine the number of factors to be extracted from the original matrix. (The number of factors to be extracted can be specified by you or you can specify the minimum eigenvalue bound). The maximum likelihood method provides a statistical basis for judging the adequacy of a model with a specified number of factors.

The unrotated factors do not generally represent useful scientific factor constructs and hence it is usually necessary to rotate. Orthogonal quartimax or varimax rotations and/or oblique rotations may be performed on a factor matrix. After rotation, a table of the variance extracted by each factor is printed along with the new factor loading matrix.

The program can graphically represent the original variables in terms of their factor loadings in a space that corresponds to the common factors. Thus, using pairs of axes, one obtains p points (where p is the number of variables) whose coordinates are factor loadings with respect to pairs of the common factors (before and after rotations).

If the raw data has been input, factor scores for each factor may be computed and stored after each rotation. These factor scores can be plotted in pairs.

Special Considerations

Factor Extraction Methods

For more information on the comparisons between the principal axes and maximum likelihood methods of factor extraction, see references 1,2 and 3.

Principal Axes Method

a. The maximum number of factors must be less than p, the number of variables in the analysis and must also be less than 15.
b. In choosing the minimum eigenvalue size for inclusion of a factor some analysts use a value around 1.00. Keep in mind that if the variables were uncorrelated, each eigenvalue would be 1.00 with the sum (total variance) equal to p.
c. The maximum number of iterations is set by default at 25. Some analysts believe that this number should be very small, say one or two.
d. The total variance is by convention, p, the number of variables in the analysis.

Maximum Likelihood Method (MLM)

a. If p is the number of variables in the analysis, then the maximum number of factors (m) which can be extracted by the MLM cannot exceed the largest integer satisfying

$$m < \tfrac{1}{2}((2p+1) - (8p+1) \uparrow .5).$$

This quantity is calculated in the program and displayed as the maximum number of factors that you may extract. See reference 11 for a more detailed discussion.

b. This method may be very time consuming. If you have a large number of variables, we suggest that you consider using the principal axes method instead.

c. This method may not converge at all. If this seems to be the case (i.e., the number of iterations and/or "tries" within an iteration is excessive), the program will allow you to stop and change to the principal axes method.

d. The chi-square statistic and hence the accuracy of the probability value depend on the number of observations being quite large. If your sample size is small you should interpret the chi-square values as only an approximation to the adequacy of the model. Some authors suggest that you should specify a fairly large value for alpha in the goodness-of-fit test, especially when the sample size is small.

Rotations

Oblique rotation schemes available in this set of programs consist of solutions generated under the oblimin criterion. A whole class of rotations may be performed, as the oblimin solution is indexed by a constant ranging between 0 and 1. The most important and generally applicable special case is bi-quartimin, which corresponds to an index value .5. Other important special cases are quartimin (index = 0) and covarimin (index = 1.0). A thorough discussion of these methods is given in (3).

Kaiser normalization will be used automatically in the program.

Output at each rotation stage consists of both primary factors and reference factors. These two types of factors are related by transformation though they are subject to different interpretations. In fact, columns of the primary factor matrix are simply multiples of the corresponding columns in the reference factor matrix. It should be noted, that since they are the elements of the primary factors (as in the orthogonal case), these elements may be larger than 1.00. It is the primary factors which are used in factor score calculations. The distinction between the aforementioned concepts is well explained in (2) and (3).

Select New Variables

After completing an analysis on certain variables and subfiles, you may wish to select other variables and/or subfiles for further analyses. You may specify the variables and subfiles you wish to investigate by choosing this option.

When you decide to select new variables, the program will go back to the beginning of the PC and FA procedures.

When entering the variable numbers, you may enter the numbers separated by commas, or by dashes when denoting consecutive variables, i.e., 1, 3, 6, 8-11 for variables 1, 3, 6, 8, 9, 10, 11.

Discussion

The purpose of this section is to reacquaint you with some of the fundamentals of principal components and factor analysis. Of course, it will not be possible to cover all of the material that would be necessary to understand all aspects of principal components and factor analysis in this section. Several of the references do have very good discussions on the basics of factor analysis and how it can be used. In particular, Sections 1.1, 1.2, and 1.3 of reference #11 have a very good discussion of the basics of Factor Analysis. In addition, reference #9 has some good material in Chapters 1, 3, 4 and 5. The other references also have some useful material.

The basic idea of multivariate statistical methods which fall into the category labeled Factor Analysis is to examine a matrix expressing the dependence structure of the response variables and to determine certain factors which have generated the dependence in these responses. We measure p variables on n individuals. These p variables frequently are interrelated, that is, they are not independent of one another. The objective of factor analysis and principal components is to find certain hidden, or latent, factors which are fewer in number than the original p variables. Ideally, the observable variables may be represented as functions of the latent factors in such a way that the original dependence structure among the responses will be generated by the new system, to some degree of accuracy. Hopefully, the number of latent variables or factors will be considerably less than p, the original number of variables. In simplest terms, the responses may be thought of as linear combinations of the latent factors, and the goal of factor analysis is to estimate the coefficients of these linear combinations.

If we are fortunate, the coefficients of the latent factors, sometimes called factor scores, will have some meaningful interpretation in terms of the original p variables. We would hope that the number of factors, or latent variables, would be considerably less than p. Ideally, two or three primary latent variables can be used in interpreting the results of the experiement. They are essentially new variables – new response variables that we can use in evaluating the results of the experiment.

This program performs a principal component analysis and factor analysis on a correlation matrix. Given the response variables $X_1, X_2, \ldots, X_p$, the technique of principal components tries to find the coefficients, say, $A_{11}, A_{21}, \ldots, A_{p1}$ such that the linear combination

$$Y_1 = A_{11}X_1 + A_{21}X_2 + \ldots + A_{p1}X_p$$

"explains" the greatest proportion of the total response variance. Having found the desired set of values, we then seek new coefficients, say, $A_{12}, A_{22}, \ldots, A_{p2}$ such that the linear combination

$$Y_2 = A_{12}X_1 + A_{22}X_2 + 4 \ldots + A_{p2}X_p$$

is uncorrelated with Y_1 and so that Y_2 explains the largest portion of the response variance remaining after Y_1 has been removed. In principal component analysis, we proceed in this manner until we have obtained $Y_1, \ldots, Y_p$. Since the Y's are chosen to be uncorrelated, their total response variance will be the same as the original $X_1, \ldots, X_p$. These linear combinations of the X's are called principal components, Y_1 being the first principal component, Y_2 being the second principal component, etc. In fact, the coefficients $A_{1j}, A_{2j}, \ldots, A_{pj}$ of the jth principal component are the elements of the eigenvector of the sample correlation matrix **R** corresponding to the jth largest eigenvalue l_j. The importance of the jth component is

measured by l_i/p. Then, if a large proportion, say 80%, of the total response variance for the X's is accounted for by a few of the Y's, we will have obtained a smaller description of the initial dependence structure. This is the main object of principal component and factor analyses – reduction of dimensionality. The program computes the principal components, eigenvalues, proportion of the total variance, and cumulative proportion of the total variance accounted for by each component.

For a study of the dependence structure, factor analysis is another technique for explaining the covariance of the responses. Principal components is simply a transformation of the responses. Factor analysis proposes a model for the responses which may be written as

$$X_1 = \lambda_{11}Y_1 + \lambda_{12}Y_2 + \ldots + \lambda_{1m}Y_m + e_1$$

$$\vdots$$

$$X_p = \lambda_{p1}Y_1 + \lambda_{p2}Y_2 + \ldots + \lambda_{pm}Y_m + e_p$$

where Y_j is called the jth common factor variable, λ_{ij} is a coefficient reflecting the importance of the jth factor for the ith response variable, and e_i is called a specific factor variable. Under this model, each response variable, X_0 is expressed as a linear combination of a few common factor variables $Y_1, \ldots, Y_m$. Let $\mathbf{F} = (\lambda_{ij})$, then $\mathbf{F}$ is the so-called factor loading matrix, the quantity

$$hi^2 = \sum_{j=1}^{m} \lambda^2_{ij}$$

is called the communality of the ith variable, and the variance of e_i is called the unique variance of the ith variable. If we replace the diagonal elements of the sample correlation matrix $\mathbf{R}$ with communalities and denote it by $\mathbf{R}^*$ then

$$\mathbf{R}^* = \mathbf{FF}'$$

This equation has been called "the fundamental factor theorem".

You can choose either the principal axes method or the maximum likelihood method to extract the initial factors. A brief comparison between these two methods can be found in reference 2. Factors which are not rotated do not generally represent useful scientific factor constructs and hence it is usually necessary to rotate. The desire for correlated (oblique) factors or uncorrelated (orthogonal) factors leads to either an oblique rotation or orthogonal rotation of the initial factor solution.

The program computes the case scores for either principal components or factors if the raw data has been input. For detailed information on the calculation and the interpretation of case scores, see Chapter 16 of reference 3.

The program also provides a graphical presentation of the initial and rotated factors.

Methods and Formulae

Correlation Matrix:

Raw Data Input:

Let the input consist of N cases with p variates per case and let $\mathbf{X} = (X_{ij})$, $i = 1, \ldots, N; j = 1, \ldots, p$, denote the data input matrix. The covariance matrix $\mathbf{S} = (s_{ij})$ is computed from

$$(N - 1)\,\mathbf{S} = \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i' - N\bar{\mathbf{x}}\bar{\mathbf{x}}'$$

where $\mathbf{X}_i' = (x_{j1}, \ldots, x_{jp})$,

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

The correlation matrix, which is used for the principal components analysis and/or factor analysis, is then given by

$\mathbf{R} = (r_{ij})$ where $r_{ij} = s_{ij}/(s_{ii}s_{jj})^{1/2}$

Covariance or Correlation Matrix Input:

Let the input consist of a matrix for p variates. For a covariance matrix, the $p(p+1)/2$ distinct elements of the matrix $\mathbf{S}$ are entered and the correlation matrix $\mathbf{R} = (r_{ij})$ is computed by

$r_{ij} = s_{ij}/(s_{ii}s_{jj})^{1/2}$

In the third method of input, the distinct elements of $\mathbf{R}$ are entered directly.

Principal Components Analysis:

The eigenvalues and corresponding eigenvectors of $\mathbf{R}$ are obtained by a variant of the QR method (see page 219 of reference 5). Let the eigenvalues of $\mathbf{R}$ be denoted by $\theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta p$ and let $\mathbf{W} = (w_{ij})$ be a pxp matrix of column eigenvectors (i.e., the jth column of $\mathbf{W}$ consists of the elements of the eigenvector corresponding to the jth eigenvalue θ_j). Then $\mathbf{W}$ is a matrix of principal components and θ_i is the variance accounted for by the ith component.

Case Scores:

For each data case a vector of component scores $\mathbf{f}$ is computed by

$\mathbf{f} = \mathbf{W}'\mathbf{z}$

where $\mathbf{W}$ is the matrix of principal components and $\mathbf{z}$ is the vector of standardized values of the variables.

Factor Extractions

Principal Axes Method:

The main diagonal elements of **R** are either unaltered or adjusted by one of the following options:

(i) squared multiple correlations on the main diagonal where r_{ii} is given by $r_{ii} = 1 - 1/r^{ii}$ and r^{ii} is the ith diagonal element of $\mathbf{R}^{-1}$. The Cholesky square root method is used to obtain $\mathbf{R}^{-1}$
(ii) maximum absolute row value among r_{ij}, $j = 1,\ldots,p$
(iii) User specified values.

The p eigenvalues and corresponding eigenvectors of **R** are obtained by the QR method. Let the eigenvalues of **R** be denoted by $\theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta p$ and the matrix of column eigenvectors be denoted by $W = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p)$. The number of factors obtained is $M = \min\{m, \#$ of θ_i such that $\theta_i > +c\}$, where M is the maximum number of factors (user specified) and c is the minimum eigenvalue for factor inclusion (also user specified). Then the jth column of the factor loading matrix $\mathbf{F} = (f_{ij})$ is $\sqrt{\theta_j}\mathbf{w}_j$. New estimates of communalities are then given by

$$\mathbf{r}_{ii} = \sum_{j=1}^{M} \mathbf{f}^2_{ij}$$

If more than one iteration is requested, the diagonal of **R** is adjusted by the new estimates of communalities and the extraction procedure is repeated. Iterations are continued until the maximum number is reached or until the maximum change in the communality estimates is less than 0.0001. If for a particular iteration any of the estimates of communalities exceed one, the process will terminate, a message will be printed, and the factor matrix for the previous iteration will be printed. Note that the number of factors may change during the iterative process.

Maximum Likelihood Method:

The Enslein procedure (see reference 13) is used to obtain the maximum likelihood solutions of the factor loading matrix **F** and the unique variance θ_{ii} of the ith variable. If k is the number of factors and

$$f_k(\Phi) = -\log \prod_{i=k+1}^{p} \Theta_i + \sum_{i=k+1}^{p} \Theta_p - (p - k)$$

where $\theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta_p$ are the eigenvalues of $\Phi^{-1/2}\mathbf{R}\Phi^{-1/2}$ and where $\Phi = \text{diag}\,(\phi_{11}, \phi_{22}, \ldots, \phi_{pp})$, the ML solution of ϕ_{ii} is the value $f\phi_{ii}$ which minimize the value of $f_k(\Phi)$. The factor loading matrix **F** is then computed by

$$\mathbf{F} = \Phi^{1/2}\mathbf{W}\,(\mathbf{H} - \mathbf{I})^{1/2}$$

where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k)$, $\mathbf{H} = \text{diag}\,(\Theta_1, \Theta_2, \Theta_k)$ and where $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ are the eigenvectors corresponding to the k largest roots. The initial estimate of $\theta_{ii} = (1 - k/2p)/r^{ii}$

where r^{ii} is the ith diagonal element of $\mathbf{R}^{-1}$. The minimization procedure of the method of Fletch and Powell is applied to the function $f_k(\Phi)$. For a detailed explanation of the computation procedure, see reference 13.

The program performs a sequence of maximum likelihood factor analyses for $k = k_1, k_1 + 1, k_1 + 2, \dots, k_2$, where k_1 is the minimum number of factors. The sequence terminates when the maximum number of factors k_2 is reached or when a proper solution has been found and is acceptable from the point of view of goodness-of-fit at a user specified level of significance. If for a particular k the solution is improper (Heywood, see reference 3), having $q < k$ of the unique variances equal to "zero", the corresponding q variables are eliminated and the partial correlation matrix $\mathbf{R}_{22x1}$ is computed as follows:

(i) Find $\mathbf{R}^{-1}$ by square root method

(ii) Delete the q columns and rows from $\mathbf{R}^{-1}$ and evaluate the inverse of the resulting matrix denoted by $\mathbf{R}_1$

(iii) $\mathbf{R}_{22x1} = \mathbf{D}^{-1/2}\mathbf{R}_1\mathbf{D}_1^{-1/2}$ where $\mathbf{D}_1$ is a diagonal matrix with the diagonal elements of $\mathbf{R}_1$

The matrix $\mathbf{R}_{22x1}$ of order $(p-q)$ is analyzed as before with the number of factors $k-q$, and the resulting solution is again examined for properness. The procedure repeats until a proper solution has been found for some $k>0$. A goodness-of-fit test is performed on this solution by computing

$$\chi^2 = [N - 1 - (2p + 5)/6 - 2k/3] \log \left[\frac{\Phi + \mathbf{FF}'}{\mathbf{R}}\right]$$

with degrees of freedom

$$v = [(p-k)^2 - p - k]/2$$

Note that $\mathbf{R}$ can be either the original correlation matrix or the partial correlation matrix, and p is the order of $\mathbf{R}$. If the computed chi-square value is greater than the tabled value with a prescribed level of significance, the value of k is increased by one and the above procedure is repeated. If the solution is acceptable, then the process terminates.

The final solution is combined with the principal components of the eliminated variables (see equations (56), (57) of reference4), if any, to give a complete solution for all the original variables.

Factor Rotation:

Orthogonal Rotation:

(i) Quartimax method: The object of the quartimax method is to determine the orthogonal transformation matrix $\mathbf{T}$ which will carry the original factor matrix $\mathbf{F}$ into a new factor matrix $\mathbf{B} = (b_{ij})$ for which

$$Q = \sum_{i=1}^{p} \sum_{j=1}^{k} b_{ij}^4$$

is a maximum. See page 298 of reference 3 for a detailed discussion.

(ii) Varimax method: The orthogonal varimax criterion requires that the final factor matrix $\mathbf{B} = (b_{ij})$ maximize the function

$$V = p\sum_{i=1}^{p}\sum_{j=1}^{k}(b_{ij}/h_i)^4 - \sum_{j=1}^{k}\left(\sum_{i=1}^{p} b_{ij}^2/h_i^2\right)^2$$

where

$$h_i^2 = \sum_{j=1}^{k} f_{ij}^2$$

the communality of the ith variable of the initial factor matrix. See page 304 of reference 3 for a detailed discussion.

Oblique Rotation:
Oblique oblimin rotation may be performed to minimize the value

$$B = \sum_{i<j=1}^{k}\left[p\sum_{l=1}^{p}(V_{li}^2/h_l^2)(V_{lj}^2/h_l^2) - \lambda\sum_{l=1}^{p}V_{li}^2/h_l^2\sum_{l=1}^{p}V_{lj}^2/h_l^2\right]$$

where

$$h_i^2 = \sum_{j=1}^{k} f_{ij}^2$$

is the communality of the ith variable of the initial factor matrix. λ is the rotation constant in the range 0 to 1. Values of λ which yield standard oblique rotations are:

(i) Quartimin: $\lambda = 0$; least oblique
(ii) Biquartimin: $\lambda = 0.5$; less oblique
(iii) Covarimin: $\lambda = 1$; most oblique

Both reference and primary factors are obtained. See page 324 of reference 3 for a detailed discussion.

Factor Scores:

Computation of factor scores begins with the calculation of a factor score coefficient matrix **C** where **C** is PXM, P is the number of variables and M the number of factors. If we let **F** be the given factor matrix (either orthogonal or oblique factors), and **R** the correlation matrix for the original data, **C** is calculated in one of two ways.

Orthogonal Factors:

$$C = R^{-1}F$$

Oblique Factors:

$$C = R^{-1}FQ$$

where F is an oblique primary factor matrix and Q is the correlation matrix of the primary factors.

Once C has been computed, the factor scores, f, for each data case are computed by

$$f = c'z$$

where z is the vector of standardized values of the variables. For detailed information on the calculation of the primary factor matrix and the Q matrix above, interpretation of the primary factors, reference structure matrix, and factor scores, see reference 3.

References

1. Enslein, K., Ralston, A., and Wilf, H. S. (eds.) (1977) Statistical Methods for Digital Computers, John Wiley & Sons, Inc., New York.
2. Gnanadesikan, R. (1977) Methods for Statistical Data Analysis of Multivariate Observations, John Wiley & Sons, New York.
3. Harman, H. H. (1967) Modern Factor Analysis, 2nd ed., University of Chicago Press, Chicago.
4. Joreskog, K. G. (1967), "Some Contributions to Maximum Likelihood Factor Analysis". Psychometrika, Vol. 32, p 443-482.
5. Martin, K. (1978) 9845B Numerical Analysis Library, Vol. 1., Hewlett-Packard Part No. 09845-10351.
6. Morrison, D. F. (1976) Multivariate Statistical Methods, 2nd ed., McGraw-Hill Book Company, New York.
7. Vecchia, D. F. Unpublished Notes for 9830A Factor Analysis.
8. Cooley, William W. and Lohnes, Paul R. (1971) Multivariate Data Analysis, John Wiley and Sons, Inc., New York.
9. Guertin, Wilson H. and Bailey, John P., J. (1970), Introduction to Modern Factor Analysis, Edwards Brothers, Inc., 1970, Ann Arbor.
10. Horst, Paul, (1965) Factor Analysis of Data Matrices, Holt, Rinehart and Winston, Inc., New York.
11. Morrison, Donald A. (1965) Multivariate Statistical Methods, Holt, Rinehart and Winston, Inc., New York.
12. Comrey, Andrew L. (1973) A First Course in Factor Analysis, Academic Press, New York.
13. Enslein, Kurt (Ralston, A. & Wilf, H. eds.) Statistical Methods for Digital Computers, Volume 4, John Wiley and Sons, Inc., New York.

Examples

Sample Problem #1

This example uses a simple artificial data set which is given below. The raw data was entered in keyboard mode. The principal component analysis was performed. Notice the "% of total variance" row corresponds to random data. Component plots of component 1 vs. component 2 and component 1 vs. component 3 were generated. Component scores were output and a plot of component scores was made, again for the same pairs of components.

Factor analysis by the principal axes method was done. Communalities were found by iteration. The iterations are not output on the printer but do appear on the CRT. The number of factors chosen to explain the variation was 3 in this example. Factor rotation plots were made for factor 1 vs. factor 2 and factor 1 vs. factor 3. An orthogonal varimax rotation was performed. The contribution of factors, % of total variance, and factor plots were output. Factor scores were also output.

Case No.	X_1	X_2	X_3	X_4	X_5
1	7	9	6	5	2
2	5	5	4	6	2
3	1	2	3	4	5
4	1	6	5	2	3
5	4	6	5	2	5
6	7	9	6	6	5
7	6	5	3	2	1
8	9	8	6	5	3
9	4	6	5	2	1
10	6	5	4	3	5
11	3	2	1	6	5
12	5	6	5	2	3
13	6	5	4	5	4
14	1	6	5	8	9
15	9	8	9	6	5
16	7	3	1	9	5
17	1	5	9	3	7
18	3	5	0	7	9
19	6	2	4	8	6
20	4	6	4	2	8

```
******************************************************************************
*                              DATA MANIPULATION                              *
******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
1                                                   Raw data
Mode number = ?
2                                                   On mass storage
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
PFACSMPB1:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES



                          SAMPLE PROBLEM #1


Data file name: PFACSMPB1:INTERNAL

Data type is:   Raw data

Number of observations:    20
Number of variables:        5


Variable names:
   1. X1
   2. X2
   3. X3
   4. X4
   5. X5

Subfiles:  NONE



SELECT ANY KEY
                                                    Press special function key labeled-LIST
Option number = ?
1                                                   List all data
Enter method for listing data:
3

                          SAMPLE PROBLEM #1

Data type is: Raw data


       Variable # 1   Variable # 2   Variable # 3   Variable # 4   Variable # 5
       (X1        )   (X2        )   (X3        )   (X4        )   (X5        )

OBS#
   1        7.00000        9.00000        6.00000        5.00000        2.00000
   2        5.00000        5.00000        4.00000        6.00000        2.00000
   3        1.00000        2.00000        3.00000        4.00000        5.00000
   4        1.00000        6.00000        5.00000        2.00000        3.00000
   5        4.00000        6.00000        5.00000        2.00000        5.00000
   6        7.00000        9.00000        6.00000        6.00000        5.00000
```

7	6.00000	5.00000	3.00000	2.00000	1.00000
8	9.00000	8.00000	6.00000	5.00000	3.00000
9	4.00000	6.00000	5.00000	2.00000	1.00000
10	6.00000	5.00000	4.00000	3.00000	5.00000
11	3.00000	2.00000	1.00000	6.00000	5.00000
12	5.00000	6.00000	5.00000	2.00000	3.00000
13	6.00000	5.00000	4.00000	5.00000	4.00000
14	1.00000	6.00000	5.00000	8.00000	9.00000
15	9.00000	8.00000	9.00000	6.00000	5.00000
16	7.00000	3.00000	1.00000	9.00000	5.00000
17	1.00000	5.00000	9.00000	3.00000	7.00000
18	3.00000	5.00000	0.00000	7.00000	9.00000
19	6.00000	2.00000	4.00000	8.00000	6.00000
20	4.00000	6.00000	4.00000	2.00000	8.00000

```
Option number = ?
0                                                      Exit list procedure
SELECT ANY KEY                                         Select special function key labeled-ADV STAT
                                                       Remove BSDM media
                                                       Insert Principal Components & Factor Analysis
Use all the variables in the analysis (YES/NO) ?       media
YES
Is the above information correct ?
YES
******************************************************************************
                  PRINCIPAL COMPONENTS AND FACTOR ANALYSIS

                                SAMPLE PROBLEM #1
******************************************************************************

---where variables to be used are :
         1. X1
         2. X2
         3. X3
         4. X4
         5. X5

                               CORRELATION MATRIX


                        X2         X3         X4         X5
X1               .4204206   .1753833   .2259743  -.3753400
X2                          .6175669  -.2043786  -.2005056
X3                                    -.2764709  -.1251464
X4                                                .3879237


Do you want to store the correlation matrix ?
NO                                                     We could store the correlation matrix for later
                                                       use, if we wished.
Enter number of desired funtion:
2                                                      Select principal component analysis
Press `CONTINUE` when ready.



                       ********************************
                       * PRINCIPAL COMPONENT ANALYSIS *
                       ********************************
Enter the option for components output(1,2,3,or 4)
1                                                      Output all principal components
```

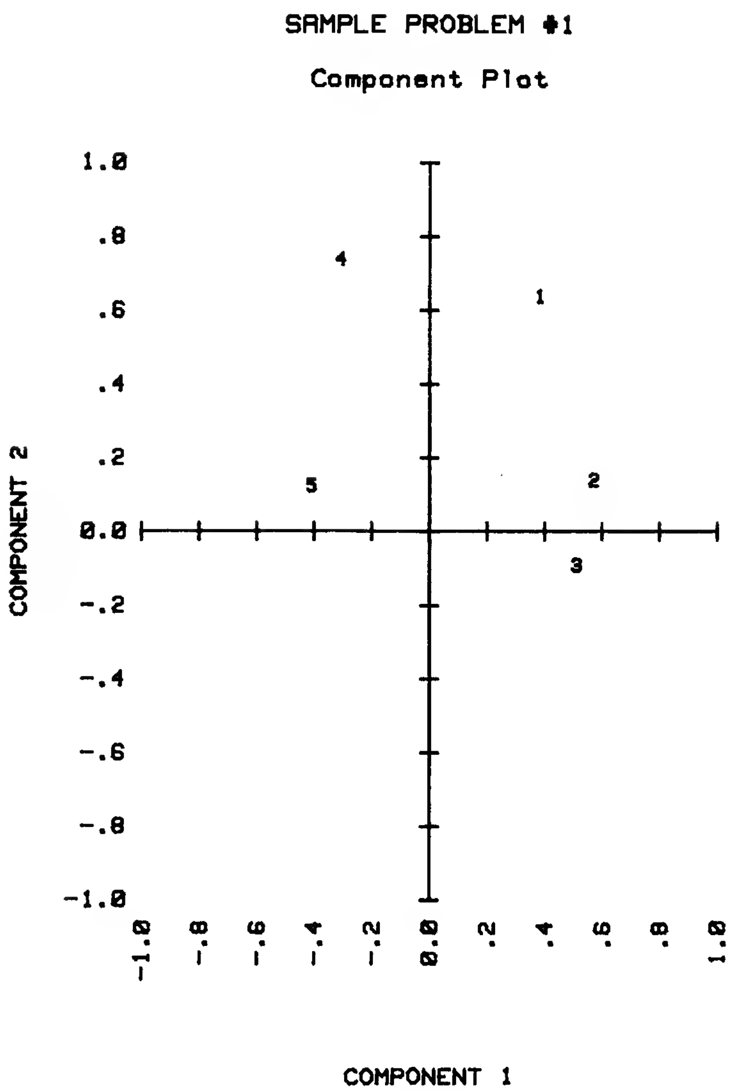
COMPONENT MATRIX

	COMPONENT				
Variable Name	1	2	3	4	5
1. X1	.383267	.637731	-.297991	-.092255	.590843
2. X2	.574271	.138684	.330269	-.584914	-.446965
3. X3	.513971	-.090709	.507831	.673708	.125823
4. X4	-.305216	.741708	.133991	.322234	-.484690
5. X5	-.407427	.125325	.725451	-.302733	.447628
Eigenvalue	2.084182	1.255467	1.046971	.363811	.249569
% of total variance	41.68365	25.10934	20.93941	7.27622	4.99139
Cumulative % variance	41.68365	66.79298	87.73240	95.00861	100.00000

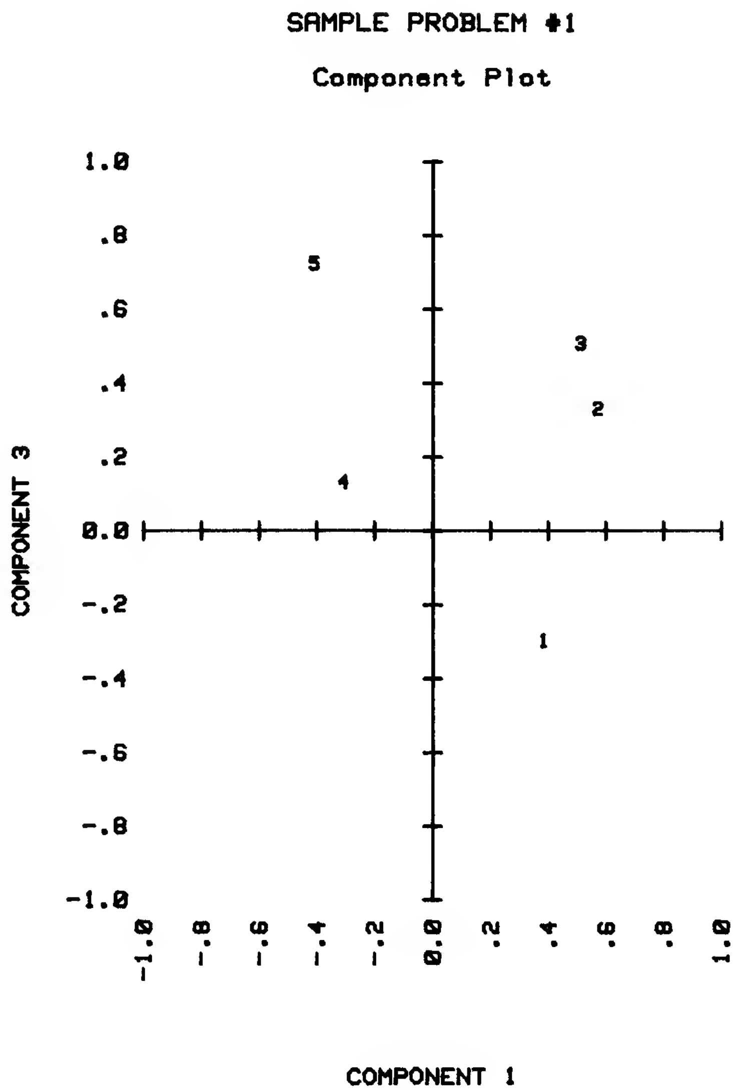
Note: First 3 principal components have Eigen values bigger than 1.0.

```
Do you wish to plot the principal components ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, HPIB bus (defaults are 7,5)?

A beep will signify the end of the plot.

Which pen number should be used ?
1
```

```
Enter the pair of component numbers which will be used in this plot ?
1,2
```



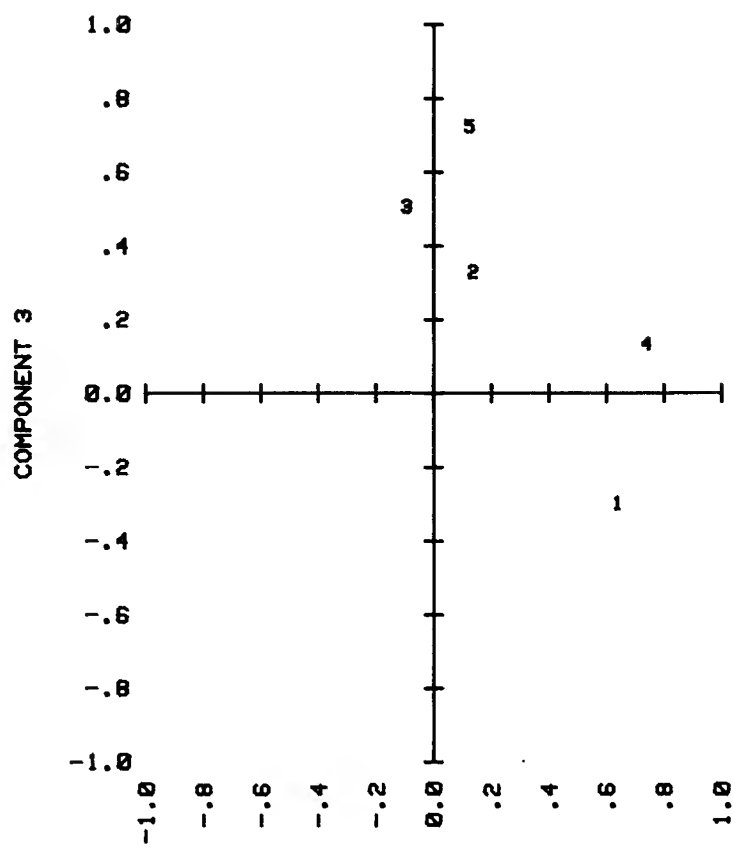
```
Plot for another two factors ?
YES
Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
1,3
```



```
Plot for another two factors ?
YES
Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
2,3
```

SAMPLE PROBLEM #1

Component Plot




```
Plot for another two factors ?
NO
Enter the option number (1,2,or 3)=
1
```

Select component scores

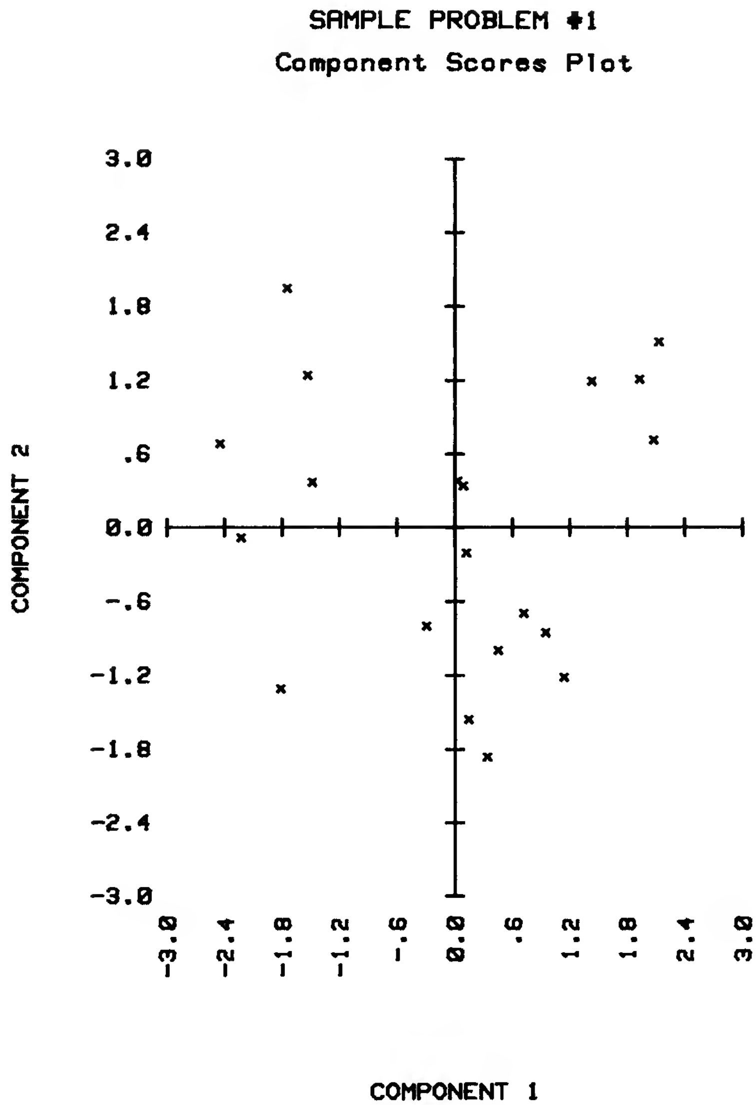
COMPONENT SCORES

	COMPONENT				
Observation #	1	2	3	4	5
1	2.07540	.71235	-.15044	-.23088	-.72271
2	.09139	.34176	-.93465	.51003	-.65276
3	-1.81738	-1.30509	-.35682	.53949	-.01545
4	.33929	-1.86345	-.00753	-.01155	-.72163

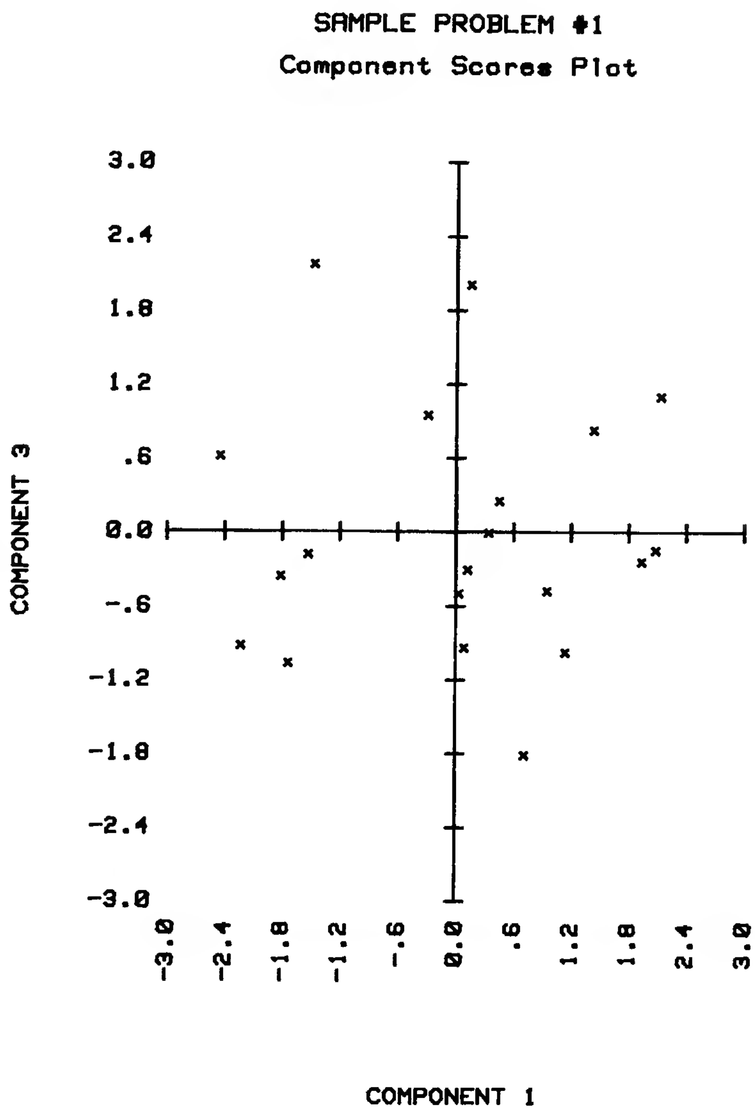
5	.44941	-1.00182	.25200	-.37656	.35664
6	1.42788	1.19038	.82627	-.47569	-.36426
7	.71513	-.69652	-1.81193	-.24860	.17100
8	1.93132	1.20276	-.23760	-.15167	.14705
9	1.13760	-1.21350	-.97337	.13479	-.39946
10	.12078	-.20532	-.30636	-.32603	.77361
11	-2.22775	-.08321	-.92196	.15357	-.07616
12	.94491	-.85573	-.47841	-.15733	.21200
13	.03008	.38027	-.49735	.07921	.16732
14	-1.48126	.36963	2.17657	.05362	-.83926
15	2.13151	1.50862	1.10035	.61701	.48187
16	-1.74141	1.94865	-1.06175	.14396	.01767
17	.14576	-1.55787	2.00757	1.07660	.26029
18	-2.44801	.68660	.61275	-1.35411	-.22557
19	-1.53268	1.24477	-.18584	1.07945	.56124
20	-.29196	-.80330	.94850	-1.05530	.86858

```
Do you wish to plot the case scores ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, HPIB bus (defaults are 7,5)?

A beep will signify the end of the plot.

Which pen number should be used ?
1
```

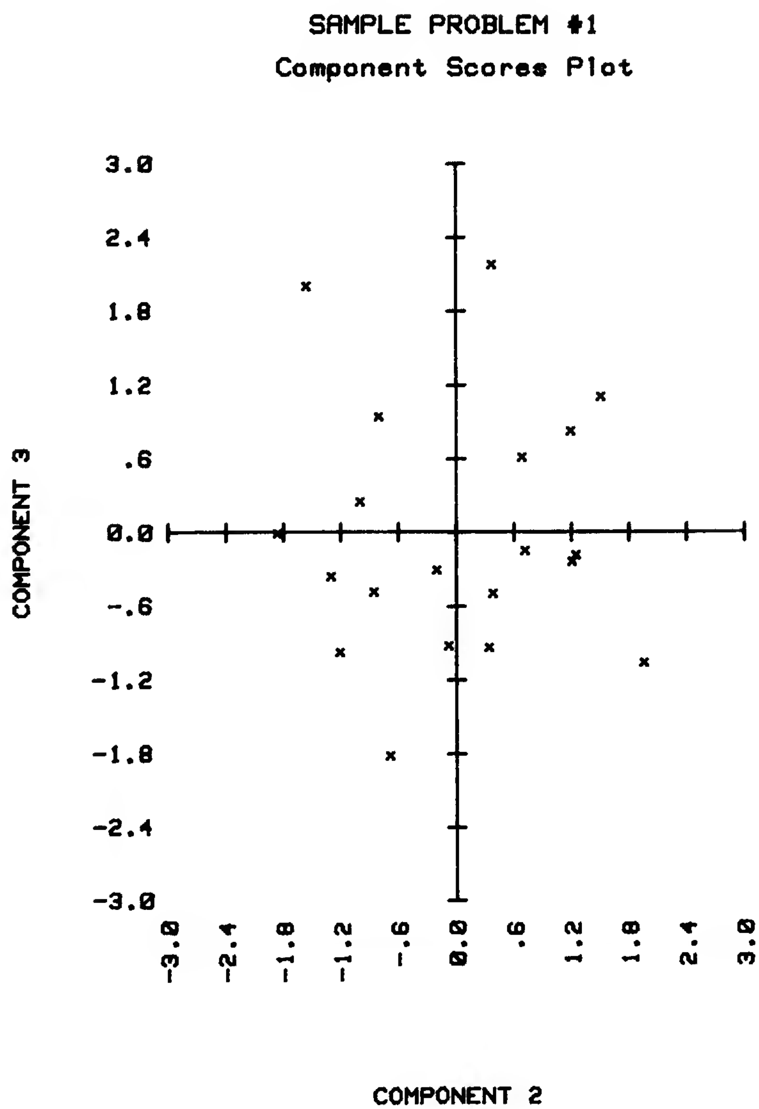
```
Enter the pair of component numbers which will be used in this plot ?
1,2
```



```
Plot for another two factors ?
YES
Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
1,3
```



```
Plot for another two factors ?
YES
Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
2,3
```




```
Plot for another two factors ?
NO
Store the principal component case scores ?
NO
Enter number of desired funtion:
3
Max. # of factors to be extracted (<= 15) :
3
```

Select factor analysis

We must specify how many factors we want to use. From the principal component analysis it appears that three might be correct.

```
*********************************************
* FACTOR ANALYSIS BY PRINCIPAL AXES METHOD *
*********************************************


A maximum of 3 factors will be extracted.
Enter Communality Estimate type (1,2,3,or 4) =
2
```

Squared multiple correlation used on the diagonal of the correlation matrix as the initial estimates.

```
COMMUNALITY ESTIMATION

Squared Multiple Correlation has been used to compute the communality estimates.

Initial Estimated Communalities of Variables :

       Variable            Communality

        1. X1                   .47407
        2. X2                   .50461
        3. X3                   .40850
        4. X4                   .42089
        5. X5                   .39380
```

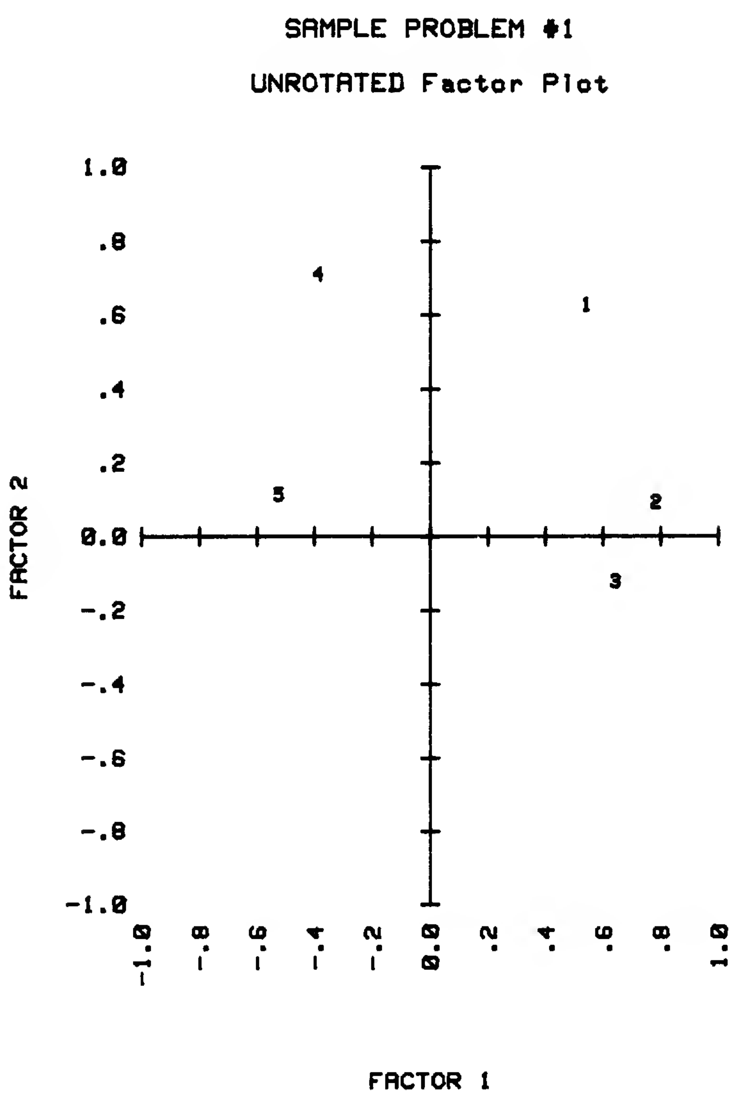
Starting values

```
Do you wish to specify a min. eigenvalue for factor inclusion ?
NO
Do you want to refine the communality estimates using iteration ?
YES
Enter the maximum # of iterations (default=25) :
5

Max. number of iterations for factor extraction = 5


Communalities of Variables after 5 iterations :

       Variable            Communality

        1. X1                   .74634
        2. X2                   .72370
        3. X3                   .57824
        4. X4                   .67900
        5. X5                   .63413
```

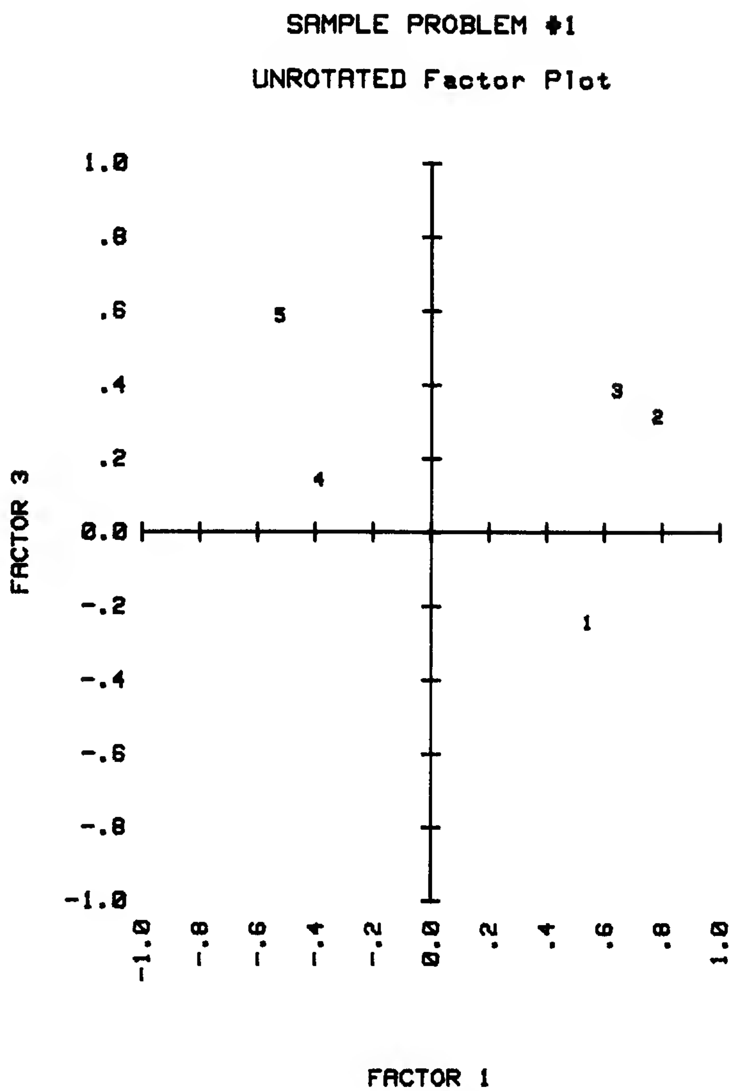
Final estimates

```
UNROTATED FACTOR MATRIX


                                              FACTOR
Variable Name        1           2           3
1. X1             .540204     .628415    -.244171
2. X2             .784661     .093539     .315046
3. X3             .644004    -.120257     .386055
4. X4            -.386153     .713522     .144134
5. X5            -.522787     .114566     .589658

Contribution
  of factor       1.74468      .94036      .67638

% of total
  Variance       34.89350    18.80713    13.52766
 Extracted
```

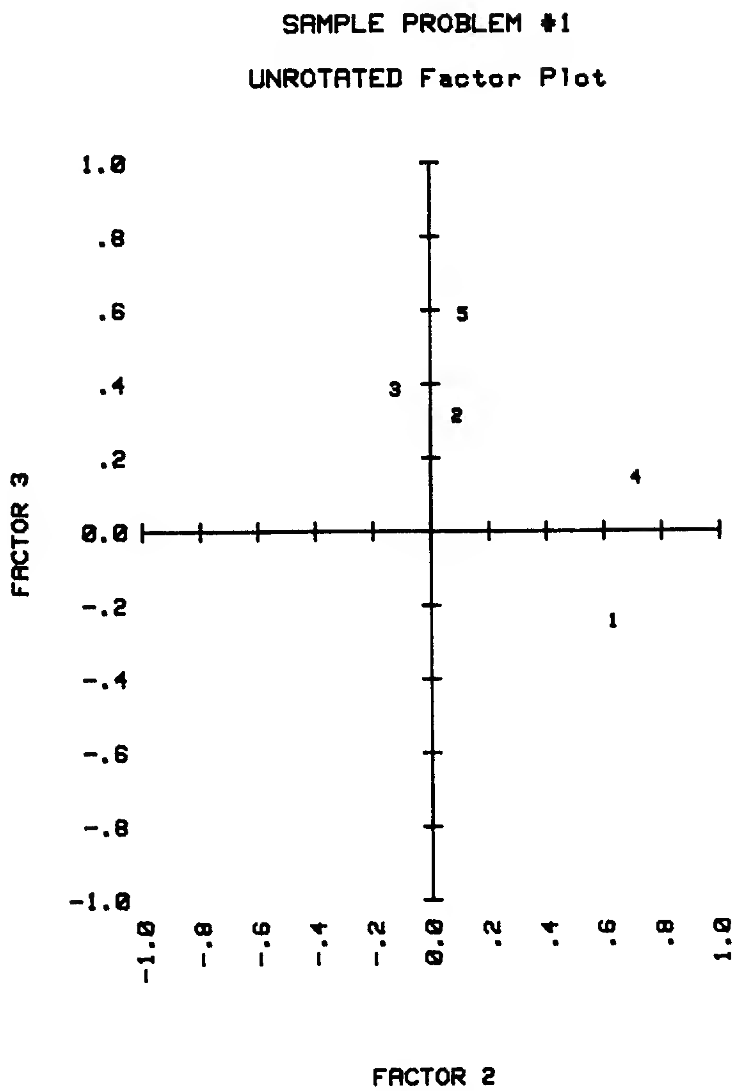
```
Do you wish to perform any factor rotations ?
YES


                              *******************
                              * FACTOR ROTATION *
                              *******************
Do you wish to plot the original factors ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter the select code, HP bus (defaults are 7,5)?

Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
1,2
A beep will signify the end of the plot.
```



```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
1,3
A beep will signify the end of the plot.
```



```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
2,3
A beep will signify the end of the plot.
```



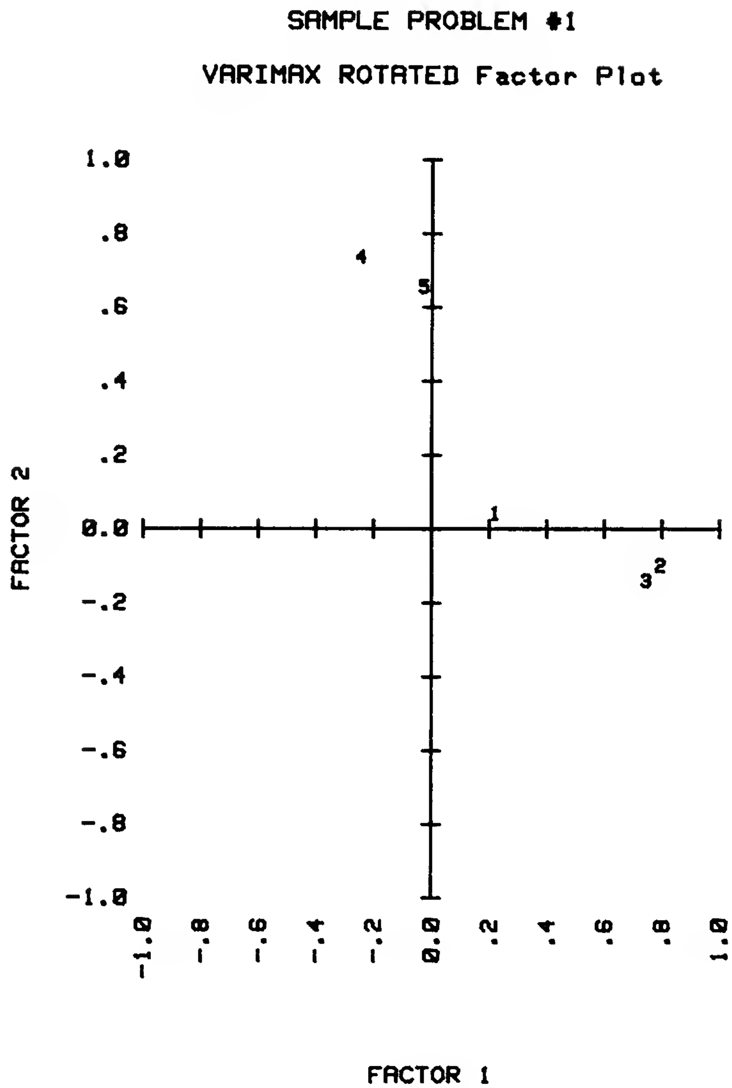

```
Plot for another two factors ?
NO
Enter the type of rotation (1 or 2) =
1                                               Orthogonal rotation
Enter the method of orthogonal rotation(1 or 2) =
1                                               Choose varimax method
```

```
                    ORTHOGONAL VARIMAX ROTATION

FACTOR MATRIX

                                        FACTOR
Variable Name          1            2            3
1. X1              .218231      .041559     -.834861
2. X2              .796148     -.099285     -.282820
3. X3              .747073     -.139647     -.024948
4. X4             -.244315      .738402     -.272169
5. X5             -.026678      .656311      .450191

Contribution
  of factor        1.30000      1.00707      1.05435

% of total
  Variance        25.99992     20.14135     21.08702
 Extracted
```

Note by the factor coefficients that factor 1 seems to be a weighted average of X2 and X3; factor 2 is a weighted average of X4 and X5, while factor 3 seems to be essentially X1 (and maybe X5).

```
Do you wish to plot the rotated factors ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter the select code, HP bus (defaults are 7,5)?

Which PEN number should be used?
1
```

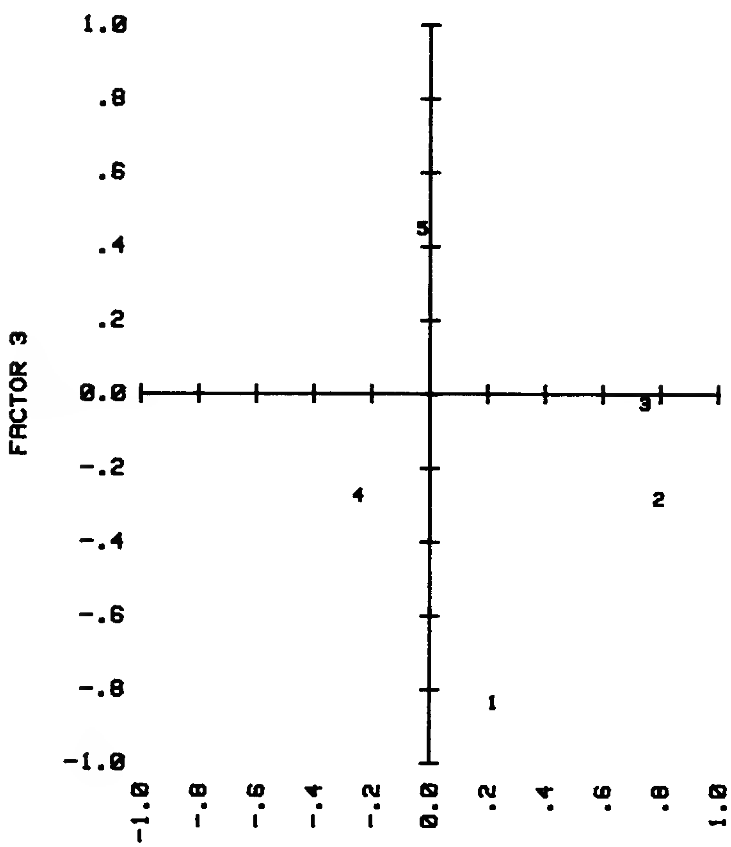
```
The pair of factor numbers used in this plot =?
1,2
A beep will signify the end of the plot.
```



```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
1,3
A beep will signify the end of the plot.
```

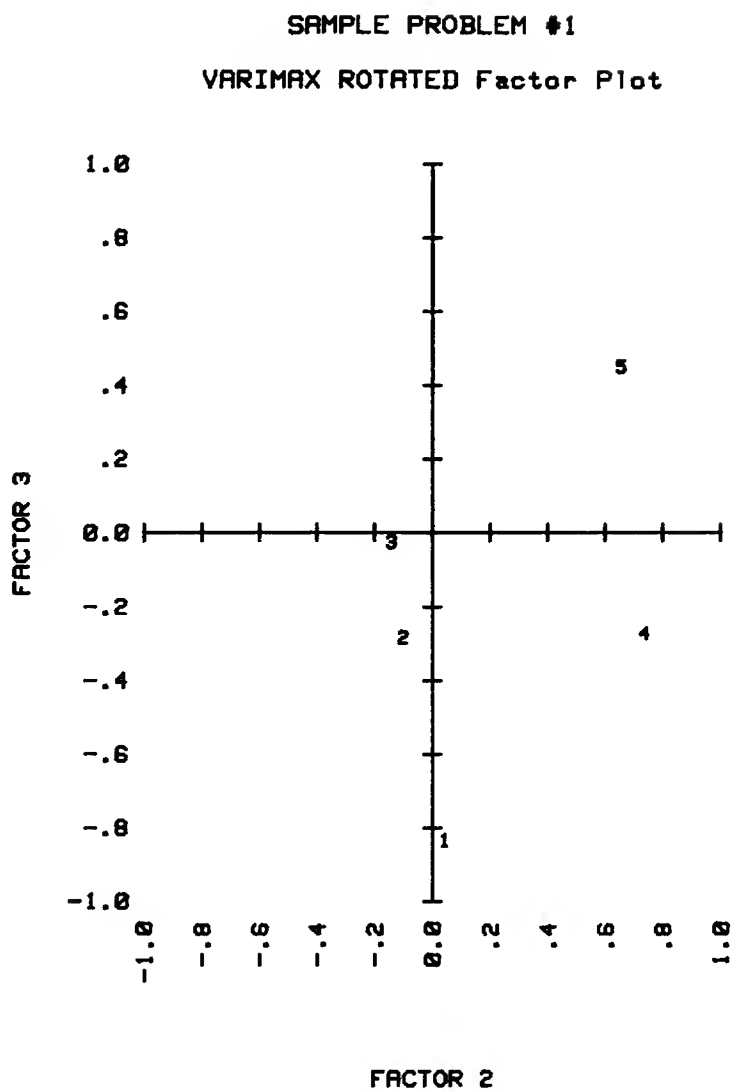
SAMPLE PROBLEM #1

VARIMAX ROTATED Factor Plot



FACTOR 1

```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
2,3
A beep will signify the end of the plot.
```




```
Plot for another two factors ?
NO
Enter the option number (1,2,or 3)=
1
```

Print out factor scores

```
FACTOR SCORE COEFFICIENTS
```

FACTOR MATRIX

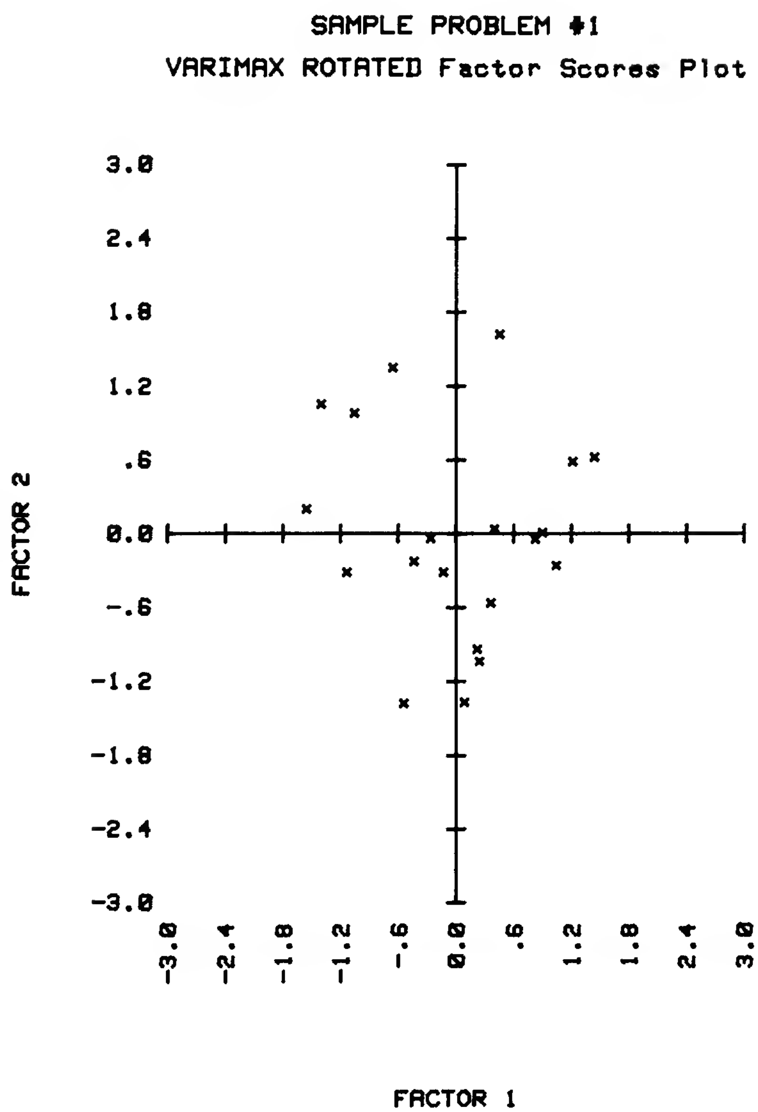
Variable Name	FACTOR 1	2	3
1. X1	-.014160	.060858	-.682742
2. X2	.576544	.074114	-.043713
3. X3	.392323	.018432	.099292
4. X4	-.078039	.558876	-.207201
5. X5	.162978	.479519	.277970

FACTOR SCORES

Observation #	FACTOR 1	2	3
1	1.03930	-.25987	-.95596
2	-.43066	-.22543	-.50906
3	-1.13434	-.30973	1.11956
4	.24275	-1.03780	1.06651
5	.36361	-.56069	.49214
6	1.21218	.58816	-.69300
7	-.54262	-1.37420	-.58291
8	.82101	-.04477	-1.35708
9	.08832	-1.37066	.02261
10	-.12901	-.31560	-.15906
11	-1.55654	.20332	.31475
12	.22038	-.94163	-.01234
13	-.26501	-.03697	-.45482
14	.45414	1.62051	1.23567
15	1.44080	.62500	-1.08097
16	-1.40375	1.05664	-1.05258
17	.89618	.00956	1.64180
18	-.65896	1.35218	.58881
19	-1.05594	.98328	-.42485
20	.39817	.03870	.80077

```
Do you wish to plot the factor scores ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter the select code, HP bus (defaults are 7,5)?

Which PEN number should be used?
1
```

```
The pair of factor numbers used in this plot =?
1,2
A beep will signify the end of the plot.
```




```
Plot for another two factors ?
NO
Do you wish to store the factor scores ?
YES
Enter a title for the new data set :
FACTOR SCORES
How many factor scores do you want to store ?
1
Name of data file :

SCORE:INTERNAL
Is data medium placed is device INTERNAL
?
YES
```

```
PROGRAM NOW STORING FACTOR SCORES
Is program medium replaced in deviceINTERNAL
?
YES


*** The 1  factor analysis scores were stored in SCORE:INTERNAL ***
Do you wish to perform another rotation ?
NO

Enter number of desired funtion:
4                                                        Return to BSDM
```

Sample Problem #2

The correlation matrix for a set of six fowl bone measurements of White Leghorn Fowl are considered. The correlation matrix is the subject of Example 7.5, page 243 of Morrison (see reference 11).

The six measurements are:

X_1 = Skull length
X_2 = Skull breadth
X_3 = Humerus
X_4 = Ulna
X_5 = Femur
X_6 = Tibia

Extraction of the principal components for the matrix reveals that 76% of the variance is explained by the first component and 88% by the first two components together. Thus, if one were interested in data reduction, it may be practical to use only the first two components (or factors).

This particular example permits an easy interpretation of the factors or components. For example, the first factor may be interpreted as a general average dimension of all bones, with the wing and leg bones receiving slightly higher loadings. Further explanation of the components may be obtained in Morrison (11).

The data was input as a correlation matrix. A principal component analysis was done and it showed that two components accounted for over 88% of the total variance. Component plots were done for component 1 vs. component 2, component 1 vs. component 3, and component 2 vs. component 3.

Factor analysis by the method of principal axes was done. Communalities were calculated. Three factors were used in the factor analysis. The first two factors accounted for over 80% of the total variance. A factor plot was done for factor 1 vs. factor 2. Then an orthogonal varimax rotation was performed. The result of the rotation and a new factor plot was output.

```
*******************************************************************************
*                            DATA MANIPULATION                                *
*******************************************************************************
Enter DATA TYPE (Press CONTINUE for RAW DATA):
3                                                   This data was stored as a correlation matrix.
Mode number = ?
2
Is data stored on program's scratch file (DATA)?
NO
Data file name = ?
BONELNGTH:INTERNAL
Was data stored by the BS&DM system ?
YES
Is data medium placed in device INTERNAL
?
YES
Is program medium placed in correct device ?
YES
```

BONE LENGTHS OF WHITE LEGHORN FOWL (MORRISON P. 243)

Data file name: BONELNGTH:INTERNAL

Data type is: Correlation matrix

Number of observations: 6
Number of variables: 6

Variable names:
1. SKULL LGTH
2. SKULL BDTH
3. HUMERUS
4. ULNA
5. FEMUR
6. TIBIA

Subfiles: NONE

SELECT ANY KEY

Press special function key labeled-LIST

BONE LENGTHS OF WHITE LEGHORN FOWL (MORRISON P. 243)

Data type is: Correlation matrix

VAR#	Variable # 1 (SKULL LGTH)	Variable # 2 (SKULL BDTH)	Variable # 3 (HUMERUS)	Variable # 4 (ULNA)	Variable # 5 (FEMUR)
1	1.00000	.58400	.61500	.60100	.57000
2	.58400	1.00000	.57600	.53000	.52600
3	.61500	.57600	1.00000	.94000	.87500
4	.60100	.53000	.94000	1.00000	.87700
5	.57000	.52600	.87500	.87700	1.00000
6	.60000	.55500	.87800	.88600	.92400

VAR#	Variable # 6 (TIBIA)
1	.60000
2	.55500
3	.87800
4	.88600
5	.92400
6	1.00000

SELECT ANY KEY

Select special function key labeled-ADV STAT
Remove BSDM media
Insert Principal Components & Factor Analysis media

Use all the variables in the analysis (YES/NO) ?
YES
Is the above information correct ?
YES

```
********************************************************************************
                  PRINCIPAL COMPONENTS AND FACTOR ANALYSIS

              BONE LENGTHS OF WHITE LEGHORN FOWL (MORRISON P. 243)
********************************************************************************

---where variables to be used are :
         1. SKULL LGTH
         2. SKULL BDTH
         3. HUMERUS
         4. ULNA
         5. FEMUR
         6. TIBIA
                                    CORRELATION MATRIX


               SKULL BDTH    HUMERUS       ULNA      FEMUR      TIBIA
SKULL LGTH       .5840000   .6150000   .6010000   .5700000   .6000000
SKULL BDTH                  .5760000   .5300000   .5260000   .5550000
HUMERUS                                .9400000   .8750000   .8780000
ULNA                                              .8770000   .8860000
FEMUR                                                        .9240000


Do you want to store the correlation matrix ?
NO

Enter number of desired funtion:
2
Press `CONTINUE` when ready.
```

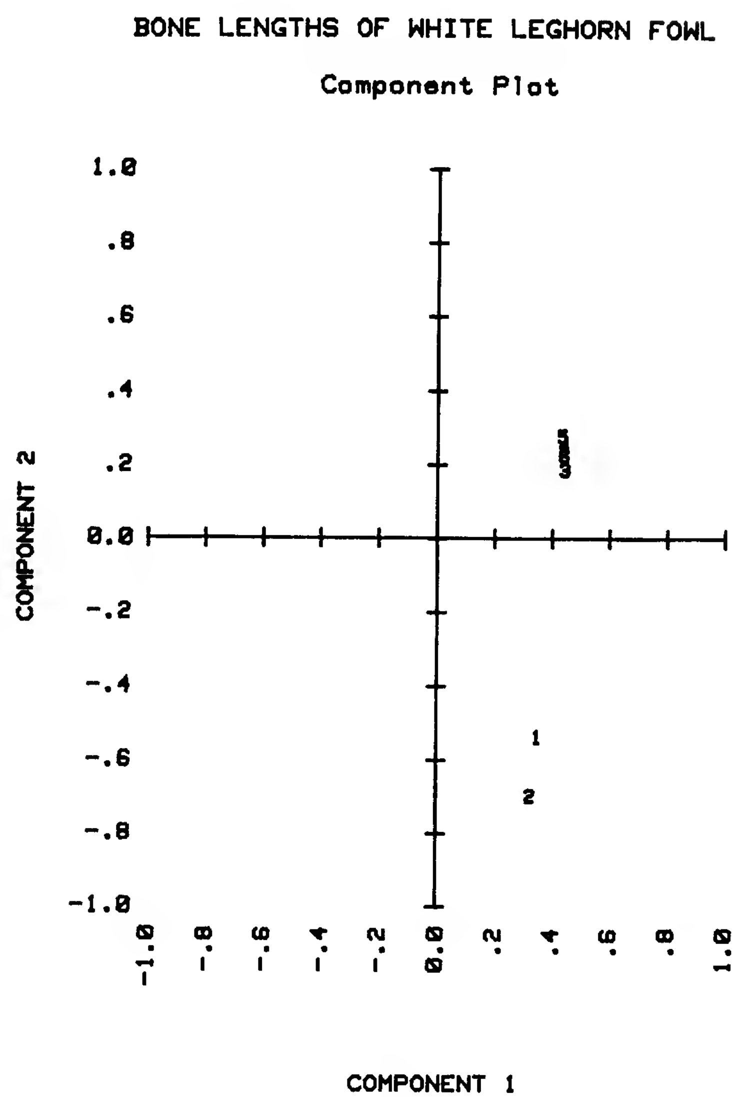
Select principal component analysis

```
                       ********************************
                       * PRINCIPAL COMPONENT ANALYSIS *
                       ********************************
Enter the option for components output(1,2,3,or 4)
1
```

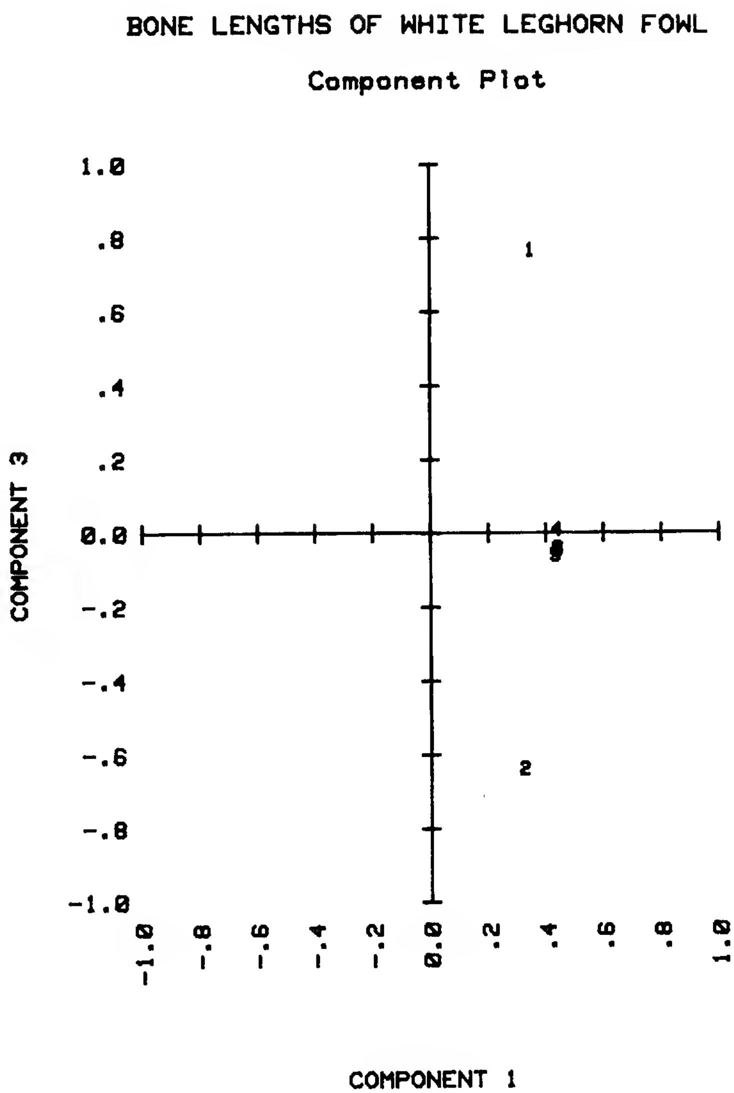
Output all the principal components

```
COMPONENT MATRIX


                            COMPONENT
Variable Name        1         2         3         4         5         6
1. SKULL LGTH    .347463  -.536959   .766673   .049099  -.027212   .002378
2. SKULL BDTH    .326404  -.696453  -.636305   .002033  -.008031   .058829
3. HUMERUS       .443411   .187321  -.040071  -.524079  -.168550  -.680900
4. ULNA          .439972   .251402   .011196  -.488769   .151309   .693763
5. FEMUR         .434532   .278188  -.059205   .514259  -.669453   .132887
6. TIBIA         .440140   .225718  -.045735   .468582   .706912  -.184237


Eigenvalue      4.567571   .714123   .412129   .173189   .075859   .057129

% of total
 variance       76.12618  11.90205   6.86882   2.88648   1.26431    .95216

Cumulative %
 variance       76.12618  88.02823  94.89705  97.78353  99.04784 100.00000
```

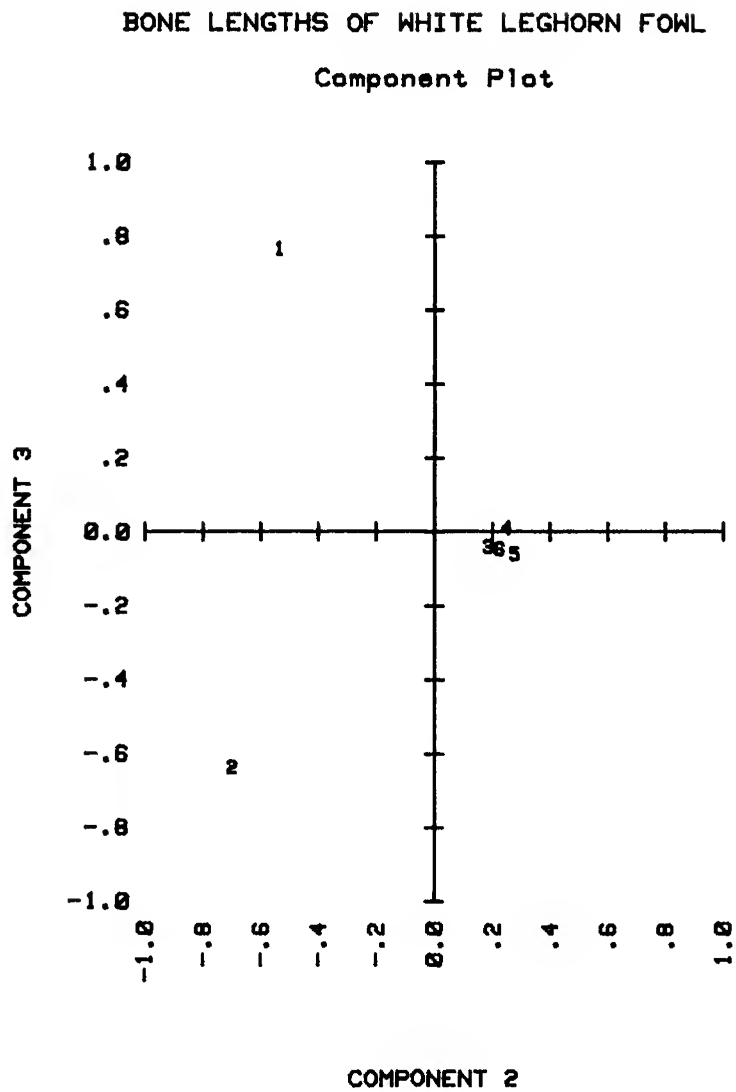
```
Do you wish to plot the principal components ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter select code, HPIB bus (defaults are 7,5)?

A beep will signify the end of the plot.

Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
1,2
```



```
Plot for another two factors ?
YES
Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
1,3
```



```
Plot for another two factors ?
YES
Which pen number should be used ?
1
Enter the pair of component numbers which will be used in this plot ?
2,3
```




```
Plot for another two factors ?
NO
Enter number of desired funtion:
3
Method for extracting factors(1 OR 2)
1
Max. # of factors to be extracted (<= 15) :
3
```

Select factor analysis

Use principal axes method

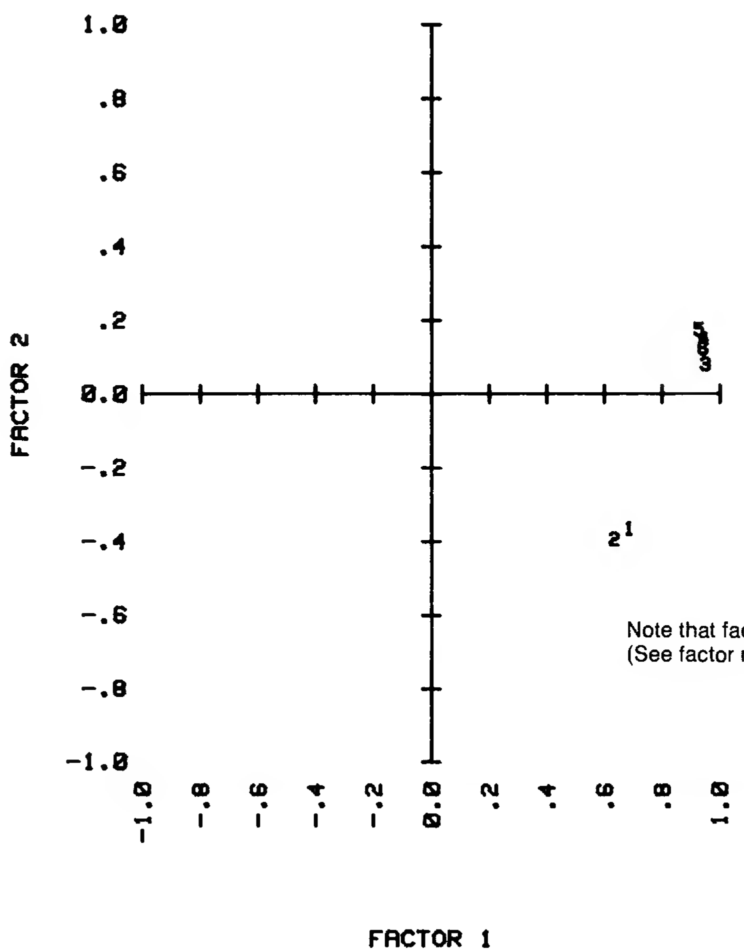
```
**********************************************
* FACTOR ANALYSIS BY PRINCIPAL AXES METHOD *
**********************************************
```

```
A maximum of 3 factors will be extracted.
Enter Communality Estimate type (1,2,3,or 4) =
2                                                    Squared multiple correlation

COMMUNALITY ESTIMATION

Squared Multiple Correlation has been used to compute the communality estimates.

Initial Estimated Communalities of Variables :

       Variable            Communality

        1. SKULL LGTH          .46814
        2. SKULL BDTH          .42741
        3. HUMERUS             .90169
        4. ULNA                .90232
        5. FEMUR               .87345
        6. TIBIA               .88329
Do you wish to specify a min. eigenvalue for factor inclusion ?
NO
Do you want to refine the communality estimates using iteration ?
YES
Enter the maximum # of iterations (default=25) :
5

Max. number of iterations for factor extraction = 5


Communalities of Variables after 5 iterations :

       Variable            Communality

        1. SKULL LGTH          .60294
        2. SKULL BDTH          .56058
        3. HUMERUS             .93835
        4. ULNA                .94385
        5. FEMUR               .91719
        6. TIBIA               .93088

UNROTATED FACTOR MATRIX


                                                FACTOR
Variable Name          1           2            3
1. SKULL LGTH       .684976    -.365703      .003721
2. SKULL BDTH       .636078    -.393993     -.027403
3. HUMERUS          .951391     .081564      .162951
4. ULNA             .945555     .150044      .165112
5. FEMUR            .928596     .176294     -.154345
6. TIBIA            .942826     .125079     -.162222

Contribution
  of factor         4.42422      .36486       .10472

% of total
  Variance         73.73696     6.08099      1.74530
 Extracted
```

```
Do you wish to perform any factor rotations ?
YES


                                   ******************
                                   * FACTOR ROTATION *
                                   ******************
Do you wish to plot the original factors ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter the select code, HP bus (defaults are 7,5)?

Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
1,2
A beep will signify the end of the plot.
```

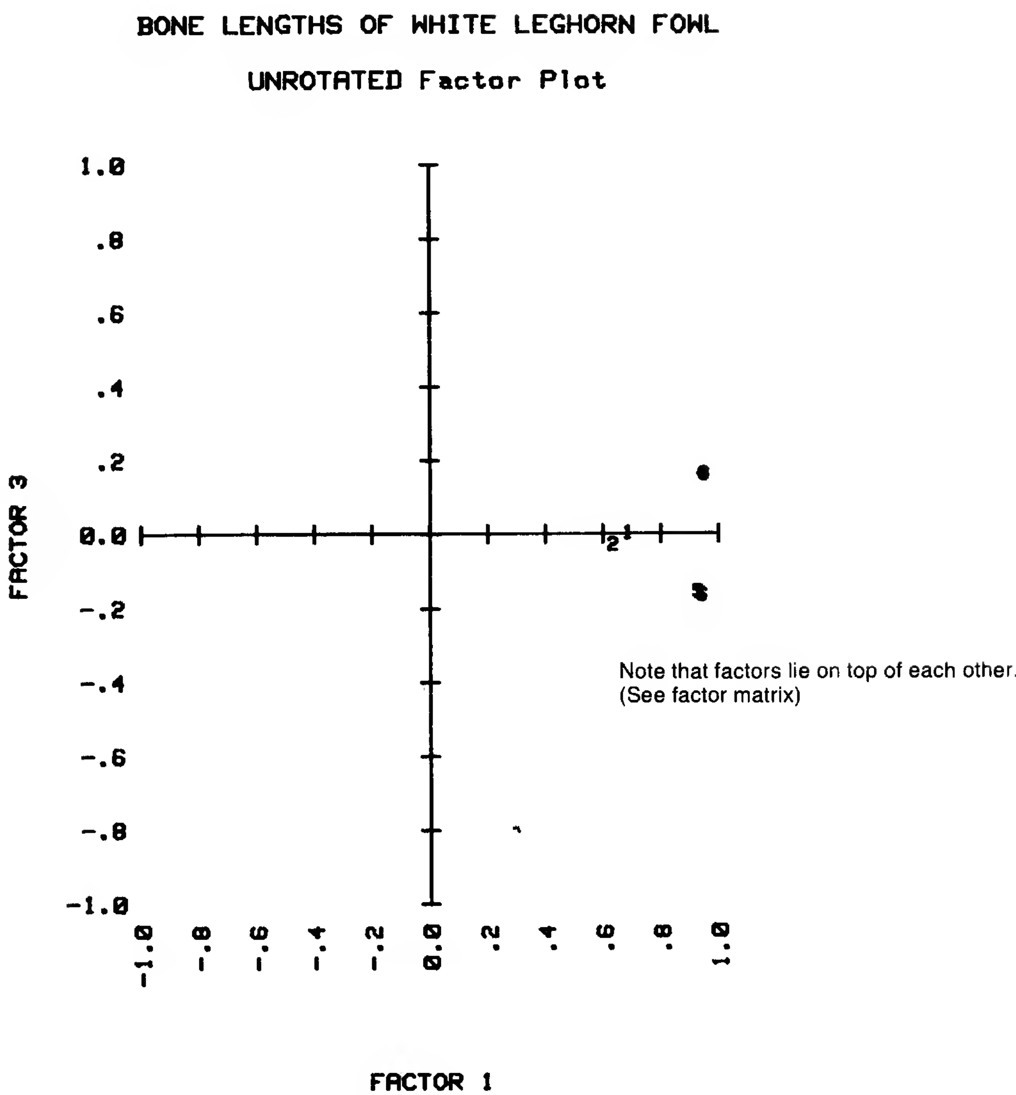
BONE LENGTHS OF WHITE LEGHORN FOWL

UNROTATED Factor Plot

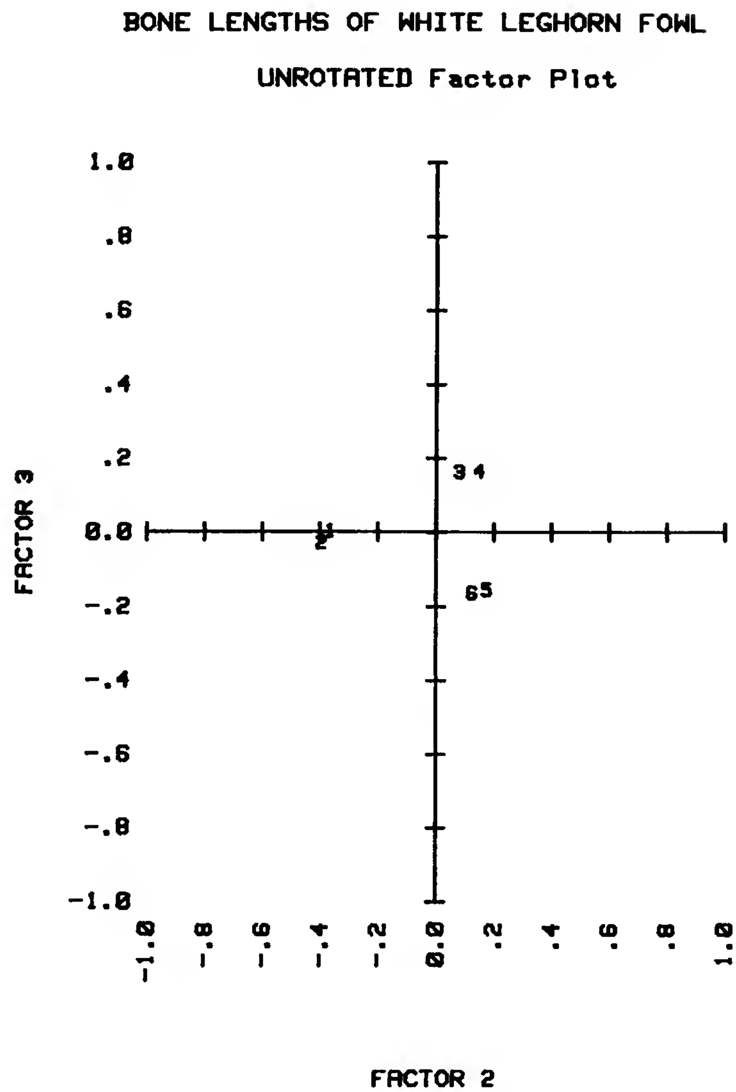


Note that factors lie on top of each other.
(See factor matrix)

```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
1,3
A beep will signify the end of the plot.
```



```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
2,3
A beep will signify the end of the plot.
```




```
Plot for another two factors ?
NO
Enter the type of rotation (1 or 2) =
1
Enter the method of orthogonal rotation(1 or 2) =
1
```

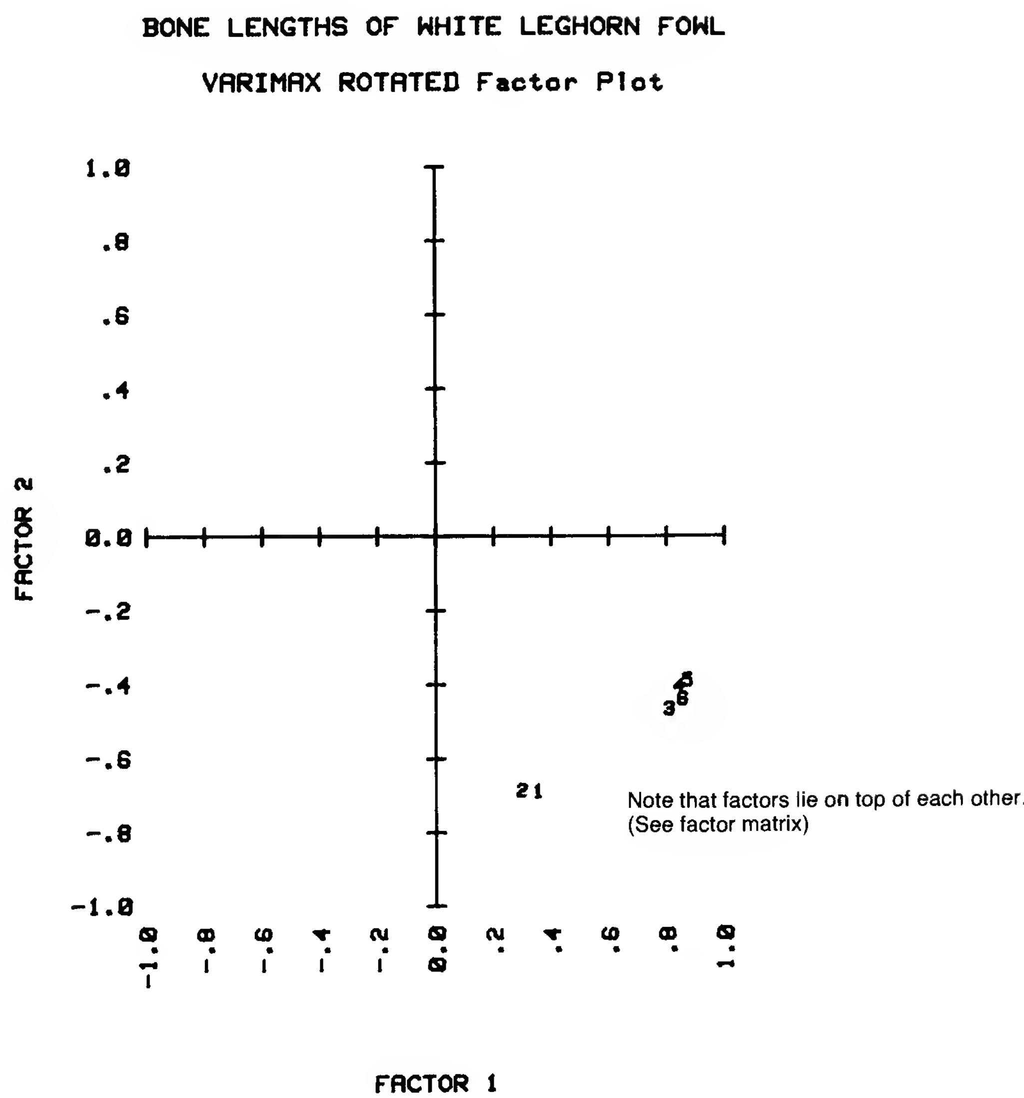
ORTHOGONAL VARIMAX ROTATION

FACTOR MATRIX

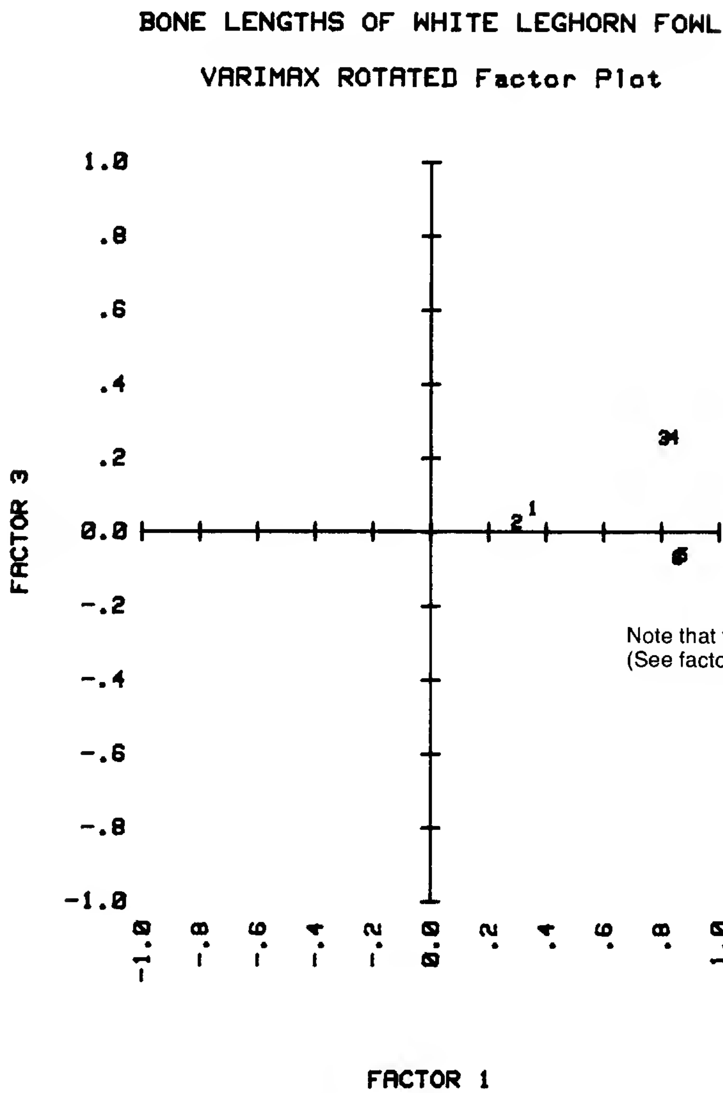
	FACTOR		
Variable Name	1	2	3
1. SKULL LGTH	.351827	-.689172	.064838
2. SKULL BDTH	.298532	-.686028	.028647
3. HUMERUS	.809812	-.465665	.256342
4. ULNA	.843788	-.405943	.259001
5. FEMUR	.873357	-.388363	-.060132
6. TIBIA	.856571	-.438891	-.067387
Contribution of factor	3.07714	1.67068	.14597
% of total Variance Extracted	51.28572	27.84462	2.43291

```
Do you wish to plot the rotated factors ?
YES
Plot on CRT ?
NO
Plotter identifier string (press CONT if 'HPGL')?

Enter the select code, HP bus (defaults are 7,5)?

Which PEN number should be used?
1
```

```
The pair of factor numbers used in this plot =?
1,2
A beep will signify the end of the plot.
```

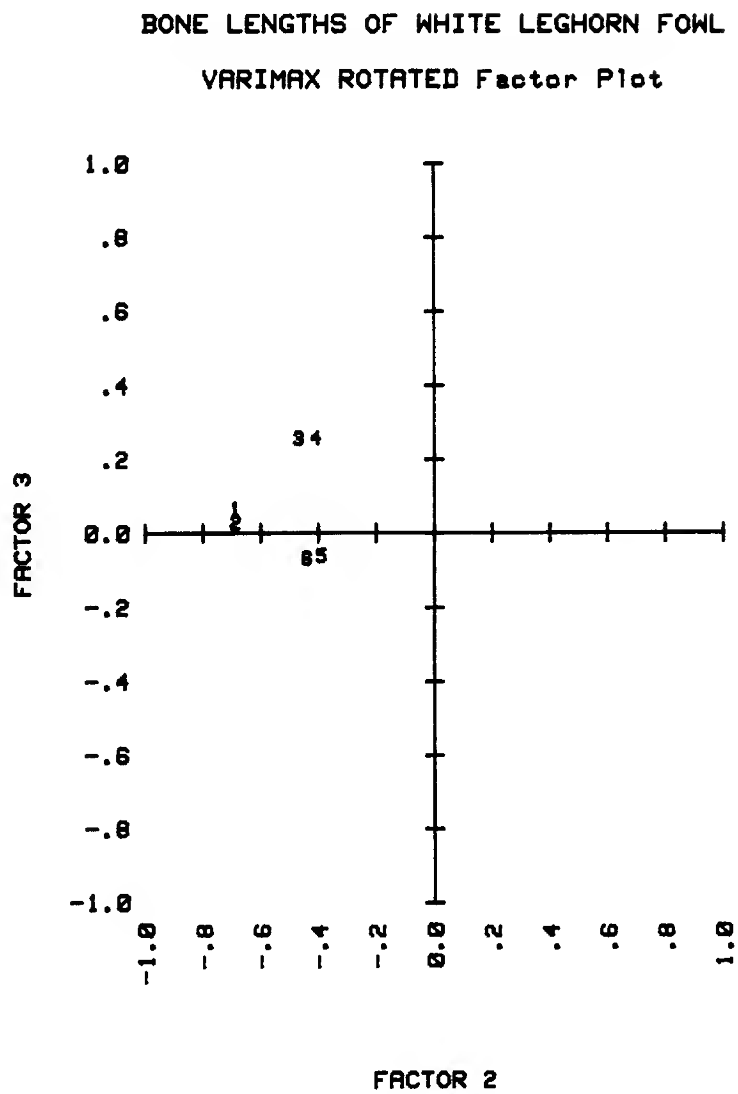


```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
 1,3
A beep will signify the end of the plot.
```



Note that factors lie on top of each other.
(See factor matrix)

```
Plot for another two factors ?
YES
Which PEN number should be used?
1
The pair of factor numbers used in this plot =?
2,3
A beep will signify the end of the plot.
```




```
Plot for another two factors ?
NO
Do you wish to perform another rotation ?
NO

Enter number of desired funtion:
4                                         Return to BSDM
```

Notes

Monte Carlo Simulations

General Information

Description

The programs in this software package are meant primarily as a library of utility routines to be combined with the user's own programs. Hence, each routine is set up as an independent, modular unit with a standard of input and output parameters. These subprograms contain no actual inputs or outputs, with the exception of error messages.

With each routine, the package provides a general-purpose front-end driver. In some cases, such as the Spectral and Run tests, the driver plus the routine make sense as a stand-alone unit. In other cases, such as the various random number deviates, the drivers are simply meant to introduce the user to the subprogram itself.

The software package **does not** establish the printers or the mass storage devices. It is the user's responsibility to select the printer and mass storage device **before** using any of these routines.

The 9826/36 operating system includes a random number generator, RND.

General Instructions

How Do I Load A Stand Alone Program?

1. Insert the program disc into the computer.
2. None of the drivers ask for the desired printer or mass storage device. This must be set by the user from the keyboard.
3. Type: LOAD "File name",10
   Press: EXECUTE.
4. At this point, appropriate inputs are requested, computations are performed, and the results are printed or saved on a mass storage device.

How Do I Add One Of The Utility Subprograms Onto My Program?

Each program file has a driver and then one or more subprograms. If you want to incorporate just one of these subprograms into your routine, how do you do it?

The entire file needs to be loaded into memory first, and then the particular subprogram needs to be saved in a temporary file. Finally, after you have written your own code, you can link the temporary file containing the desired subprogram on after your code.

1. Insert the program cartridge or disc into the computer.
2. Type: LOAD "File name"
   Press: EXECUTE
3. After the program has been loaded,
   Type: EDIT
   Press: EXECUTE
4. At this point, the screen looks as follows:

```
10   Beginning of driver program.
20        .
          .          Driver program
          .
     END
          .
          .
          .
100   SUB Sub_to_be_linked
          .
          .
          .
      SUBEND
```

5. If subprogram Sub_to_be_linked is the one desired and it goes from line 100 to line 500, then
   Type: SAVE "TEMP",100,500
   Press: EXECUTE.
6. Type: SCRATCH A
   Press: EXECUTE.
7. After you enter your program into memory, for this example assume that the last line of your code is line 2500. Then
   Type: GET "TEMP",2510
   Press: EXECUTE.
8. The desired subprogram is then linked on behind your routine.

Special Considerations

1. All the programs in this package have been set up using the random number generator RND. This may be replaced by the super random generator contained in RSUPER.
2. You now have two different random number generators at your disposal.

   RND: a randomly generated generator. (See the section further on in General Information for more details.)

   RSUPER: a combination generator. (See "RSUPER" for further details.)

It is strongly suggested that any serious Monte Carlo simulation should be run with both of these generators.

3. This package is meant to provide a set of subprogram utilities which you can combine to meet your particular needs. Each utility may be viewed as an independent modular unit. This allows you to combine these building blocks into your own program.
4. In order to get a feel for how each utility works and, in the case of the various generators, how much confidence you can place in them, driver routines have been provided. So, it is suggested that you first use these driver programs as is, and then later adapt them to your particular need.
5. In order to allow you the most flexibility, no references are made to printers or mass storage devices. Hence, to have a particular program run from a floppy disk in the internal disc drive and have all information printed on the CRT, you would type in the following before running your program:

   1. a. Type: MASS STORAGE IS ":INTERNAL"
      b. Press: EXECUTE
   2. a. Type: PRINTER IS 1
      b. Press: EXECUTE
6. Each of the driver programs for the random deviates allows you to:

   1. generate a set of random numbers to be printed or saved on a mass storage device.

      or
   2. get a feeling for the quality of the generator by running through some randomly generated tests.

7. There may be occasions where you will not have enough memory to store all the random numbers you would like to have. A number of possible tricks are available to you:

   a. Presently all deviates are set up in full precision arrays. Can you store the deviates in an integer? Where a full precision array requires 8 bytes per number, an integer only requires 2. Care must be taken here to dimension your array using an INTEGER statement rather than a DIM. Also, the parameters in the SUB statement must be changed to INTEGER.
   b. Can you generate and use the random numbers in a partitioned fashion? For example, generate 1000 deviates, use them; generate 1000 more, use them; etc.
   c. If b is not possible, can you make use of your mass storage device to recall the deviates as you need them? For example:

      i. generate 1000 deviates; store them; generate 1000 more, store them; etc.
      ii. bring first 1000 deviates into memory; use them; bring them 1000 in, use them; etc.

8. Entering a value of 1 for the printer's select code automatically causes the program to skip over the question requesting the printer's bus address.
9. If you choose to check through some examples of random data sets produced by one of the generators, default values are supplied for the parameters. For example, you may see a prompt such as:

   ```
   # OF RANDOM DEVIATES IN EACH SET?
   100
   ```

   If the default number, 100, is acceptable to you simply press CONTINUE and 100 deviates will be generated in each set. If you wish to have a different number generated, edit the number in the response line before pressing CONTINUE.
10. If you store a set of random numbers produced by one of the generators, the data set may be read into a statistical data base created by Basic Statistics and Data Manipulation (BSDM) and then accessed by any other statistics routine.

    To access the data using BSDM, remember that the data was not stored by BSDM. Thus, you will need to supply a name for the data set, a variable name, number of observations, etc.

9826/36 Random Number Generator: RND

This generator uses a standard "multiplicative congruential generator". In this generator, a starting value called the seed is multiplied by a positive integer constant, and the result is taken modulus M.

$$X_{(i+1)} = A * X_i \text{ Mod } M$$

The algorithm used in the RND has a starting seed of 37480660. This seed may be set by the program to any new value by using the RANDOMIZE statement.

In this routine, the value A = 16 807, is used for the multiplier. The modulus $M = 2^{31} - 1$. The exact steps used in the algorithm are presented below.

The algorithm below is the one used to generate the next random number in a sequence from the previous one (i.e., the seed) using RND:

1. Multiply the current seed by 16 807.
2. Take the result of Step 1 Modulus M.
3. Save result of Step 2 as the new seed.
4. Convert the result of Step 2 to a number between 0 and 1.
   (Divide by $2^{31} - 1$).
5. Go to Step 1.

References

1. Camp, Warren V. and Lewis, T.G., "Implementing a Pseudo-Random Number Generator on a Minicomputer", IEEE Transactions on Software Engineering, May, 1977.
2. Knuth, Donald E., The Art of Computer Programming, Volume 2: Seminumerical Algorithms, Addision-Wesley, Reading, Mass., 1969.
3. Learmonth, J. and Lewis, P.A.W., "Naval Postgraduate School Random Number Generator Package LLRANDOM", Naval Postgraduate School, Monterey, Calif., 1973.
4. Learmonth, J. and Lewis, P.A.W., "Statistical Tests of Some Widely Used and Recently Proposed Uniform Random Number Generators", Naval Postgraduate School, Monterey, Calif., 1973.
5. MacLauren, M.D. and Marsaglia, G., "Uniform Random Number Generators", JACM 12, Jan. 1965, p. 83-89.
6. Marsaglia, G. and Bray, T.A., "One-line Random Number Generators and Their Use in Combinations", CACM, Vol. II, 1968, p. 757-759.
7. Musyck, E., "Search For a Perfect Generator of Random Numbers", Studiecentrum Voor Kernenergie, E. Plaskylaan 144, Brussels 4, Belgium, January, 1977.
8. Reddy, Y.V., "PL/I Process Generators", SIMULETTER, Vol. III, Oct. 1976, p. 25-29.
9. Wheeler, Robert E.,"Random Variable Generators", SIMULETTER,Vol. III, Oct. 1976, p. 16-22.

Random Number Generators

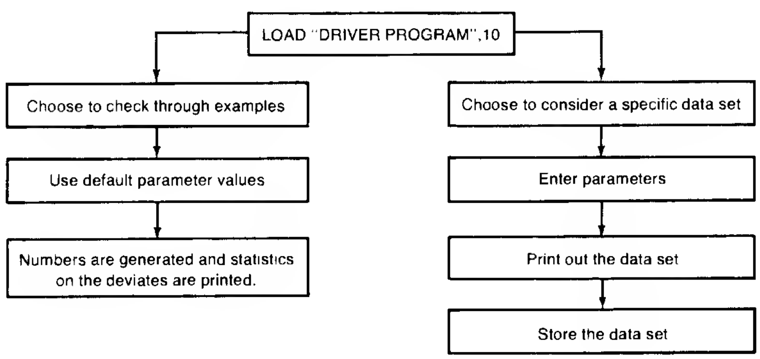
Object of Program

Subprograms with optional drivers are provided to generate random deviates on some standard statistical distributions.

The subprograms have been set up as independent modules. Hence, it is quite simple to use these routines in your own programs. Choose values for the required input parameters, call the subprogram and the resulting outputs are returned to you. See the General Information section if this manual for detailed instructions.

Optional drivers have also been set up for your use. In general, the drivers: i) allow you to directly generate a set of deviates to be printed or saved on a mass storage device; and ii) provide the ability to check out the particular generator through the use of some standard tests in order to get a feel for the quality of the deviates produced.

Typical Program Flow



(RBETA) Random Numbers Generated from a Beta Distribution

Description

Given a Beta distribution with V1 and V2 degrees of freedom, respectively, this subprogram generates a set of random deviates. The probability density function is:

f(x) = [x ↑ (V1/2 − 1)] [(1 − x) ↑ (V2/2 − 1)] / [B(V1/2,V2/2)]
for $0 \leq x \leq 1$, where B(*,*) is the beta function.

File Name

"RBETA"

Calling Syntax

CALL Random_beta (N,V1,V2,X(*))

Input Parameters

N	number of deviates desired.
V1, V2	degrees of freedom on the Beta distribution.

Output Parameters

X(*)	array of dimension (1:N) containing the N deviates.

Algorithm

This routine generates deviates for the beta distribution with v1, v2 degrees of freedom. The method used is valid for both integer and non-integer v1 and v2:

1. Generate uniform random deviates u1 and u2.
2. Set y1 = u1 ↑ (2/v1); y2 = u2 ↑ (2/v2), repeating this process until finding y1 + y2< = 1.
3. Then x = y1/(y1 + y2).

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2 Seminumerical Algorithms, Reading, Mass.: Addision-Wesley, 1969, p. 115.

(RBINOM) Random Integers Generated From a Binomial Distribution (T,P)

Description

Given that some event occurs with probability P and that we carry out T independent trials, this subprogram generates a set of integers with the binomial distribution (T,P). The probability density function is:

$f(x) = \binom{T}{x}$ [P ↑ x] [(1 − P) ↑ (T − x)]
For $x = 0, 1, \ldots, T$.

File Name

"RBINOM"

Calling Syntax

CALL Random_binomial (N,P,T,X(*))

Input Parameters

N	number of deviates
P	probability of the event occurring.
T	number of independent trials.

Output Parameters

X(*)	array of dimension (1:N) containing integers randomly generated for the number of occurrences.

Algorithm

Given T and P:

1. Set Sum = 0.
2. For I = 1 to T.
3. Generate a uniform random deviate U.
4. If U < = P then Sum = Sum + 1.
5. Next I.
6. The binomial deviate is equal to Sum.

Reference

1. Reddy, Y.V., "PL/I Process Generators", SIMULETTER, Vol III, Oct. 1976, p. 25-26.

(RCHISQ) Random Numbers From a Chi-square Distribution

Description

Given the number of degrees of freedom and the number of deviates desired, this subprogram generates a set of random numbers with the Chi-square distribution. The probability density function is:

f(x) =[.5 ↑ (v/2)][x ↑ (v/2-1)][exp(−.5x)]/[G(v/2)] for x > 0, where v is the degrees of freedom and G(*) is the gamma function.

File Name

"RCHISQ"

Calling Syntax

CALL Random_chi_sq(N,V,X(*))

Input Parameters

N number of deviates desired.

V degrees of freedom.

Output Parameters

X(*) array of dimension (1:N) containing the N deviates.

Algorithm

This utility generates random deviates for the Chi-square distribution with v degrees of freedom.

For each deviate, if v = 2*k, where k is an integer

set x = 2*(y1 + y2+...+yk) where the y's are independent random variables with the exponential distribution, each with mean = 1.

If v = 2*k+1,

set x = 2*(y1 + y2+...+yk)+z ↑ 2 where the y's are as before, and z is a random variable independent of the y's, with the normal distribution (mean = , standard deviation = 1).

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2 Seminumerical Algorithms. Reading, Mass: Addison-Wesley, 1969, p. 115.

(REXPON)
Random Numbers From an Exponential Distribution

Description
Given a mean, which you supply, this subprogram generates a set of exponential deviates. The probability density function is:

$f(x) = [\exp(-x/\mu)]/\mu$
for $x>0$, where μ is the mean of the distribution = Mμ.

File Name
"REXPON"

Calling Syntax
CALL Random_expon (N,Mu,X(*))

Input Parameters
N	number of deviates desired.
Mu	mean of the distribution.

Output Parameters
X(*)	array of dimension (1:N) containing the N deviates.

Algorithm
This routine uses the random minimization method (due to George Marsaglia) to compute an exponentially distributed variable without using the logarithm subroutine. Although this routine takes slightly more space, it is much faster than the traditional algorithm.

Reference
1. Knuth, Donald E., The Art of Computer Programming, Volume 2 Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 114.

(RF)
Random Numbers Generated From an F-Distribution

Description

Given an F-distribution (variance-ratio distribution) with V1 and V2 being the numerator and denominator degrees of freedom, respectively, this subprogram generates a set of corresponding random deviates. The probability density function is:

$$f(x) = \frac{[G(V1/2+V2/2)][(V1/V2) \uparrow V1/2][x \uparrow (V1/2-1)]}{G(V1/2)G(V2/2)[(1+(V1/V2)x) \uparrow (V1/2+V2/2)]}$$

for x>0, V1 and V2 positive integers.

File Name
"RF"

Calling Syntax
CALL Random_f(N,V1,V2,X(*))

Input Parameters
N number of deviates desired.

V1, V2 degrees of freedom on the F-distribution.

Output Parameters
X(*) array of dimension (1:N) containing the N random numbers.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2 Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 116.

(RGAMM1)
Random Integers Generated From a Gamma (Alpha) Distribution

Description

This subprogram generates a set of Gamma (Alpha) deviates. The probability density function is:

$$f(x) = [(x) \uparrow (\text{Alpha} - 1))(\exp (- x)]/G(\text{Alpha})$$

where Alpha>0 is the distribution parameter and G(*) is the gamma function.

File Name

"RGAMM1"

Calling Syntax

CALL Random_gamma1 (N,Alpha,X(*))

Input Parameters

N	number of random numbers desired.
Alpha	Gamma parameter.

Output Parameters

X(*)	array of dimension (1:N) containing numbers randomly generated with the given Gamma distribution.

(RGAMM2) Random Numbers Generated From a Gamma (A,B) Distribution

Description

This subprogram generates a set of Gamma (A,B) random deviates. The probability density function is:

f(x) = [x ↑ (B − 1)][exp (− x/A)] / [G(B) A ↑ B]
for x, A and B>0, where G(*) is the gamma function.

File Name

"RGAMM2"

Calling Syntax

CALL Random_gamma2 (N,A,B,X(*))

Input Parameters

N — number of random deviates desired.

A,B — Gamma parameters, B must be an integer.

Output Parameters

X(*) — array of dimension (1:N) containing deviates randomly generated with the Gamma distribution.

Algorithm

1. Given Gamma parameters A and B, generate B independent exponential deviates with mean = A.
2. The corresponding Gamma deviate is equal to the sum of the B exponential deviates.

(RGEOM)
Random Integers Generated From a Geometric Distribution

Description

Given that a certain event occurs with probability P, this subprogram generates N random integers with the appropriate Geometric distribution; that is, each random integer represents the number of individual trials needed until the given event first occurs (or between occurrences of the event). The probability density function is:

$f(x) = P(1-P) \uparrow (x-1)$
for $x = 1,2\dots$.

File Name
"RGEOM" **Calling Syntax**
Call Random_geom (N,P,Integer(*))

Input Parameters

N	number of random integers desired.
P	probability of a given event occurring.

Output Parameters

Integer(*)	array of dimension (1:N) containing integers randomly generated for the number of independent trials needed until the given event occurs.

Algorithm

The probability of the event first occurring on the Rth trial is $P*(1-P) \uparrow (R-1)$.

A convenient way to generate a variable with this distribution when P is small, is to set R = the least integer function of $[\ln(U)/\ln(1-P)]$ where U is a uniformly generated random number.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithnms, Reading, Mass.: Addison-Wesley, p. 116.

(RLNORM)
Random Lognormal Deviates

Description

This subprogram generates a set of random deviates such that the natural logarithm of the deviates follows a normal distribution with mean = Mu and standard deviation = Sigma. The probability density function is:

f(x) = [exp(− .5[(ln x − Mu)/Sigma] ↑ 2)]/[x((2*PI) ↑ .5)*Sigma]

File Name
"RLNORM"

Calling Syntax
CALL Random_lognorm (N,Mu,Sigma,X(*))

Input Parameters

N	number of deviates desired.
Mu	mean of the associated normal distribution.
Sigma	standard deviation of the associated normal distribution.

Output Parameters

X(*)	array of dimension (1:N) containing the N lognormal deviates.

Algorithm

1. Let S = log[(Sigma ↑ 2)/(Mu ↑ 2) + 1].
2. Let U = log (Mu) − 0.5*S.
3. Generate a normal deviate A, with mean = U and standard deviation = Square Root of (S).
4. Then the lognormal deviate is equal to exp (A).

Reference

1. Reddy, Y.V., "PL/I Process Generators", SIMULETTER, Vol. III, Oct., 1976, p. 27.

(RNEGBI)
Random Numbers Generated From a Negative Binomial Distribution

Description

This subprogram generates a set of Negative Binomial random deviates, that is, each random integer represents the number of trials needed until a given event occurs R times. The probability density function is:

$$f(x) = \binom{x-1}{R-1} \ (P \uparrow R)\ (\ (1-P) \uparrow (x-R)\ )$$

for $0 \leq P \leq 1$, and $x = 1,2\ldots$.

File Name
"RNEGBI"

Calling Syntax
CALL Random_neg_bin (N,R,P,X(*))

Input Parameters

N	number of random integers desired.
R	failure value.
P	probability.

Algorithm

1. Given parameters R and P, generate R random geometric deviates with parameter P.
2. The corresponding Negative Binomial Deviate is equal to the sum of the R geometric deviates.

Reference

1. Wheeler, R.E., "Random Variable Generators", SIMULETTER, Vol. IV, April, 1973, p. 22.

(RNORM)
Normal Random Deviates With Mean = 0 And Standard Deviation = 1

Description

This subprogram calculates an even number of normally distributed variables with mean = 0 and standard deviation = 1. The probability density function is:

f(x) = [exp(−.5(x ↑ 2))] / [(2*PI) ↑ .5]

File Name

"RNORM"

Calling Syntax

CALL Random_normal (N,X(*))

Input Parameters

N number of normal deviates desired. N must be even.

Output Parameters

X(*) array of dimension (1:N) containing the N normal deviates.

Algorithm

This utility generates random deviates for the normal distribution with mean = 0 and standard deviation = 1. An adapted form of the Polar Method is used. (See Reference 1.)

Special Considerations

1. Due to the nature of the algorithm used, this routine generates an even number of normal deviates. If an odd number is requested, an error message is printed and the routine has to be re-entered again.
2. This method is rather slow, but it has essentially perfect accuracy and takes a minimum of storage space.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 104.

(RNORM1)
Normal Random Deviates With Specified Mean and Standard Deviation

Description

This subprogram generates a set of normal random deviates with mean = Mu and standard deviation = Sigma. The probability density function is:

f(x) = exp[− (x − Mu)^2/(2*Sigma ↑ 2)]/[(2*PI) ↑ .5*Sigma]
where Sigma >0.

File Name

"RNORM1"

Calling Syntax

CALL Random_normal1 (N,Mu,Sigma,X(*))

Input Parameters

N	number of deviates desired
Mu	assume a normal distribution with mean = Mu.
Sigma	assume a normal distribution with Standard Deviation = Sigma.

Output Parameters

X(*)	array of dimension (1:N) containing the N normal deviates.

Algorithm

Given a mean = u and standard
deviation = s,

1. Generate a deviate x with a normal distribution with mean 0 and standard deviation = 1.
2. Then y = u + s * x.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 113.

(RNORM2)
Dependent Normally Distributed Random Variables (Bivariate Normal Deviates)

Description

This subprogram generates two dependent random variables which have a bivariate normal distribution with marginal means = Mu1,Mu2, marginal standard deviations = Sigma1, Sigma2, and Correlation Coefficient = Rho.

File Name
"RNORM2"

Calling Syntax
CALL Random_normal2 (Mu1,Mu2,Sigma1,Sigma2,Rho,X1(*),X2(*))

Input Parameters

Mu1, Mu2	marginal means.
Sigma1, Sigma2	marginal standard deviations.
Rho	marginal correlation coefficient.

Output Parameters

X1(*), X2(*)	two vectors of dependent normally distributed random variables.

Algorithm

If x1 and x2 are independent normal deviates with mean = 0 and standard deviation = 1, and if

$$y1 = Mu1 + Sigma1*x1, \text{ and } y2 = Mu2 + Sigma2*(Rho*x1 + \sqrt{1 - Rho \uparrow 2}*x2)$$

then y1 and y2 are dependent random variables, normally distributed with means Mu1, Mu2 and standard deviations Sigma1 and Sigma2, and with correlation coefficient Rho.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 113.

(RPAR1)
Random Pareto Generator Of The First Kind

Description

This program generates sets of random Pareto deviates of the first kind. The probability density function is defined as follows:

f(x) = [N*A ↑ N]/x ↑ (N+1) for x>A

File Name

"RPAR1"

Calling Syntax

CALL Random_pareto1 (Number A,N,X(*))

Input Parameters

Number	number of random deviates desired.
A,N	Pareto parameters.

Output Parameters

X(*)	array of dimension (1:N) containing N Pareto deviates of the first kind.

Algorithm

1. Given parameters A and N, generate a uniform deviate U.
2. Then the Pareto deviate is equal to: A/(1 − U) ↑ (1/N).

(RPAR2)
Random Pareto Generator Of The Second Kind

Description

This program generates sets of random Pareto deviates of the second kind. The probability density function is defined as follows:

f(x) = [N*B ↑ N] / [B + x] ↑ (N + 1) for x>0.

File Name
"RPAR2"

Calling Syntax
CALL Random_pareto2 (Number B,N,X(*))

Input Parameters

Number	number of random deviates desired.
B,N	Pareto parameters.

Output Parameters

X(*)	array of dimension (1:N) containing N Pareto deviates of the second kind.

Algorithm

1. Given parameters B and N, generate a uniform deviate U.
2. Then the Pareto deviate is equal to: B/(1 – U) ↑ (1/N) – B.

(RPOISS)
Random Integers Generated From A Poisson Distribution

Description

This subprogram generates a set of Poisson deviates with a specified mean. The probability density function is:

$f(x) = [\exp(-Mu)\ (Mu \uparrow x)] / x!$
for $x = 0,1,\ldots$, where Mu is the mean of the distribution, and $Mu>0$

File Name

"RPOISS"

Calling Syntax

CALL Random_poisson (N,Mu,X(*))

Input Parameters

N	number of random integers desired.
Mu	mean of the Poisson distribution.

Output Parameters

X(*)	array of dimension (1:N) containing integers randomly generated with the given Poisson distribution.

Algorithm

Given a mean of the distribution Mu,

1. Set: $P = \exp(-Mu)$
   $N = 0$
   $Q = 1$
2. Generate a random variable U, uniformly distributed between 0 and 1.
3. Set: $Q = Q*U$
4. If $Q>P$, then set $N = N + 1$ and return to step 2.
   Else, terminate the algorithm with output N.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 116.

(RSPHER) Random Points on an M-dimensional Sphere of Radius One

Description

This subprogram generates a set of random points on an M-dimensional sphere of radius one.

File Name

"RSPHER"

Calling Syntax

CALL Random_sphere (N,M,X)*))

Input Parameters

N	number of random points desired.
M	number of dimensions of the sphere.

Output Parameters

X(*)	array of dimension (1:N) containing the N random points.

Algorithm

1. Let X1, X2 . . . ., Xm be independent normal deviates (means = 0, standard deviation = 1).
2. Let R = SQR(X1 ↑ 2 + X2 ↑ 2 + ... + Xm ↑ 2).
3. Then the point (X1/R,X2/R,...,Xm/R) is a random point on the M dimensional sphere of radius one.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2 Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 116.

(RSUPER)
Super Uniform Random Number Generator

Description

Given methods for generating two random sequences, this schuffling algorithm successfully outputs the terms of a 'considerably more random' sequence. This routine uses RND twice to generate 'super' random numbers and, due to the slow execution speed, should be used only in cases where no regular random number generator will do. The probability density function is:

$f(x) = 1$
for $0 \leq x \leq 1$

File Name

"RSUPER"

Calling Syntax

CALL Random_super (N,X(*))

Input Parameters

N number of random deviates desired.

Output Parameters

X(*) array of dimension (1:N) containing N uniformly generated random numbers on the range (0,1).

Algorithm

This method has been suggested by Bays and Durham in (Ref. 1). Given methods for generating two pseudo-random sequences xn and yn, this routine will output terms of a 'considerably more random' sequence.

A temporary table V(1:107) is used in the generation of sequence yn.

1. Fill table V with the first 107 elements of sequence Xn.
2. Set X,Y equal to the next numbers of the sequences Xn,Yn, respectively.
3. Set J = INT(101*Y + 1)
4. Output V(J) and set V(J) = X.
   Go to step 2.

In our routine, both sequences Xn and Yn are generated using RND.

Knuth contends that the sequence obtained by applying this algorithm will satisfy virtually anyone's requirements for randomness in a computer-generated sequence.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Vol. II. Seminumerical Algorithms, Second Edition, Reading, Mass.: Addison-Wesley, 1969, 1981.

Special Considerations

1. As a result of our own tests, this generator comes highly recommended. It performed extremely well on all of our tests of randomness. In terms of execution speed and storage space, it is approximately three times as slow as RND alone, plus it requires an extra 856 or so bytes for storage of the temporary array.
2. In using this routine, it is suggested that as many random deviates be generated on one call as is possible. Each time the subprogram is entered, 107 new table values are created.
3. If you are interested in repeatability of an experiment, remember that initial seeds must be set for RND (using RANDOMIZE).
4. If you plan on calling this routine a large number of times, a significant amount of time would be saved if the table V is set up once in your calling routine and then passed as an additional parameter to Random_super. This will avoid the overhead of redoing this table each time you enter the routine.

(RT)
Random Numbers Generated From A T-Distribution

Description

This subprogram generates a set of random deviates for a T-distribution with V degrees of freedom. The probability density function is:

$$f(x) = G(\ (V+1)/2)/[G(V/2)\ (\ (V*PI)\ \uparrow\ .5)\ (\ (1+(\chi\ \uparrow\ 2)/V(\ \uparrow\ (V+1)/z]$$

for $V = 1,2,\ldots$

File Name

"RT"

Calling Syntax

CALL Random_t (N,V,X(*))

Input Parameters

N number of random deviates desired.

V degrees of freedom.

Output Parameters

X(*) array of dimension (1:N) containing the N random deviates.

Algorithm

1. Let y1 be a normal deviate. (mean = 0, standard deviation = 1)
2. Let y2 be independent of y1, having the Chi-square distribution with v degrees of freedom.
3. Then x = y1/(SQR(y2/v)) is independent, having the T distribution with v degrees of freedom.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 116.

(RT1EXT)
Random Type I Extreme-Value Generator

Description

This program generates sets of random Type I Extreme-Value deviates. The cumulative distribution function is defined as follows:

f(x) = exp(− exp[− Alpha*(x − Mu)])

File Name
"RT1EXT"

Calling Syntax
CALL Random_type1ext (Number,Alpha,Mu,X(*)).

Input Parametes

Number	number of random deviates desired.
Alpha Mu	Type I parameters.

Output Parameters

X(*)	array of dimension (1:N) containing N Type I deviates.

Algorithm

1. Given parameters Alpha and Mu, generate a uniform deviate U.
2. Then the Type II deviate is equal to: − log[− log(U)]/Alpha + Mu.

(RT2EXT)
Random Type II Extreme-Value Generator

Description

This program generates sets of random Type II Extreme-Value deviates. The cumulative distribution function is defined as follows:

F(x) = exp[– (V/x) ↑ K]

File Name

"RT2EXT"

Calling Syntax

CALL Random_type2ext (Number,V,K,X(*))

Input Parameters

Number	number of random deviates desired.
V,K	Type II parameters.

Output Parameters

X(*)	array of dimension (1:N) containing N Type II deviates.

Algorithm

1. Given parameters V and K, generate a uniform deviate U.
2. Then the Type II deviate is equal to: V*[– log(U)] ↑ (– 1/K).

(RUNIF)
Uniform Random Number Generator

Description

This program generates sets of uniform random numbers. The probability density function is:

$$f(x) = 1$$

for $0 \leq x \leq 1$

Calling Syntax

CALL Random_uniform (N,X(*))

Input Parameters

N number of random deviates desired.

Output Parameters

X(*) array of dimension (1:N) containing N uniformly generated random numbers on the range (0,1).

(RWEIBU)
Random Integers Generated From a Weibull Distribution

Description

This subprogram generates a set of Weibull deviates. The cumulative distribution function is:

$$F(x) = 1 - \exp[-(x \uparrow (\text{Beta}))/\text{Alpha}]$$

File Name

"RWEIBU"

Calling Syntax

CALL Random_weibull (N,Alpha,Beta,X(*))

Input Parameters

N	number of random deviates desired.
Alpha,Beta	Weibull parameters.

Output Parameters

X(*)	array of dimension (1:N) containing deviates randomly generated with the given Weibull distribution.

Reference

1. Wheeler, R.E., "Random Variable Generators", SIMULETTER, Vol. IV, April 1973, p. 22.

Tests for Randomness

Object of Programs

A standard set of statistical tests for randomness is provided. These tests are designed as independent subprograms with optional drivers. These driver programs have been set up to test the binary random number generator RND for randomness. The aim here is twofold: i) to actually allow you to check the randomness of RND; and ii) to show you how a typical test might be set up.

(TCHISQ)
Chi-square Test

Description

This subprogram performs a Chi-Square test on a set of observations placed in a set of categories with given probabilities.

File Name
"TCHISQ"

Calling Syntax
Call Chi_sq_test (N,Cats,Prob(*),Obs(*),V,P)

Input Parameters

N	number of observations. This should be at least 5*Cats, but preferably much larger, for a valid test.
Cats	number of categories.
Prob(*)	array of dimension (1:Cats) containing the probabilities of any event occurring in a particular category. Care must be taken to insure that no probability value is too small.
Obs(*)	array of dimension (1:Cats) containing the number of observations occurring in each category.

Output Parameters

V	Chi-square statistic. V is expected to have the Chi-square distribution with (Cats − 1) degrees of freedom.
P	right-tailed probability; Prob (X>V).

Special Considerations

1. The Chi-square method can only be used with sets of independent observations.
2. The proper choice of N is somewhat obscure. Large values of N will tend to smooth out 'locally' non-random behavior, that is, blocks of numbers with a strong bias followed by blocks of numbers with the opposite bias. But, N should be large enough so that each of the expected values N*Prob> = 5 for the probability associated with each category. Preferably, N should be taken much larger than this. So, the method should probably be used with a number of different values of N.
3. From the Chi-square formula, we can see that a very small probability value would severely influence the Chi-square statistic. Hence, it is suggested that categories with very small probabilities be grouped together into one larger category.
4. You must supply the routine with the number of categories into which the data is to be partitioned. For example, to check the randomness of the first digit, ten categories will be sufficient. To check the first two digits, 100 categories are recommended.

Algorithm

A fairly large number, N, of independent observations is made. We count the number of observations falling into each of K categories, and compute the quantity.

$$V = (1 / N) \sum_{i=1}^{K} (\, (\text{observed}(I) \uparrow 2)/\text{Prob}(I)\,) - N$$

In the associated driver program, the right-tailed probability P(X>V) is then calculated using (K − 1) as the number of degrees of freedom.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 35-40.

(TKS)
Kolmogorov-Smirnov Test

Description

Given a continuous cumulative distribution function F(X), this subprogram calculates the standard Kolmogorov-Smirnov statistics of maximum deviation.

File Name

"TKS"

Calling Syntax

Call K_s_test (N,Knp,Knn)

Input Parameters

N	number of observations

The distribution function F(X) must be provided as an in-line function to the subprogram.

Output Parameters

Knp	positive K-S statistic.
Knn	negative K-S statistic.

Algorithm

Given a distribution function F(x) = probability that (X< = x) for a random variable X, the statistics Knp (Kn positive) and Knn (Kn negative) can be obtained as follows:

1. Obtain the observations x1,x2,..., xn.
2. Sort the observations: x1< = x2< = ...< = xn.
3. Knp = SQR(n)* maximum of [j/n − F(xj)] where 1< = j< = n.
   Knn = SQR(n) * maximum of [F(xj) − (j − 1)/n] where 1 < = j< = n.

Special Considerations

1. The method used in the driver program (using several tests for moderately sized N, then combining the observations later in another K-S test), tends to detect both local and global nonrandom behavior.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 41-48.

(TMAXT)
Maximum of T Test

Description

This routine generates groups of uniform random numbers, finds the maximum of each group and then applies the Kolmogorov-Smirnov test to the resulting set of numbers.

File Name

"TMAXT"

Calling Syntax

CALL Max_of_t (N,T,Knp,Knn)

Input Parameters

N	number of groups to be tested.
T	size of each group.

Output Parameters

Knp	positive Kolmogorov-Smirnov statistic.
Knn	negative Kolmogorov-Smirnov statistic.

Algorithm

For 0< = j<n, let Vj = max(Utj, Utj + 1, ..., Utj + t − 1) where the U's are uniformly distributed random numbers.

Now apply the Kolmogorov-Smirnov test to the sequence V0, V1, ..., Vn − 1, with the distribution function F(x) = x ↑ t, (0< = x< = 1).

Reference

1. Knuth, Donald E., The Art of Computer Programming, Vol. II, Seminumerical Algorithms, Readinbg, Mass.: Addison-Wesley, 1969, p. 64.

(TPOKER)
Modified Poker Test

Description

This subprogram calculates the number of distinct values in a given set of observations. A Chi-square test is then applied to the set of data.

File Name

"TPOKER"

Calling Syntax

CALL Poker_test (K,N,Digits,V,P)

Input Parameters

K — number of possible different digits in a set. The degrees of freedom is then $(K-1)$. A reasonable number here is 5.

N — number of test sets to be used. N should be at least 5*(K-1), but preferably much larger, for a valid Chi-square test.

Digits — range on the allowed digits, [0,Digits-1]; 13 or 10 would be reasonable values here.

Output Parameters

V — Chi-square statistic. V is expected to have the Chi-square distribution with (K-1) degrees of freedom.

P — right-tailed probability; Prob $(X>V)$.

Algorithm

In general, we look at n groups of k successive numbers. We count the number of k-tuples with r different values. For example, generate 1000 groups of 5 successive numbers, where the numbers range from 1 to 13. How many sets have all 5 numbers different? How many have 4 different? How many 3? 2? 1?

A Chi-square test is then made, using the probability.

$$P(r) = d*(d-1)*...*(d-r+1)/(d \uparrow k)*S(k,r)$$

where d is the number of possible digits considered and S(k,r) is the standard Sterling number of k,r.

Special Considerations

You will be required to enter a starting and ending value for the number of groups desired, as well as the increment between values. At each value, three independent tests are run.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 57-58.

(TRUNS)
Runs Test

Description

This subprogram sets up N random numbers and calculates the number of ascending or descending runs in the sequence. A special Chi-square statistic is then produced.

File Name

"TRUNS"

Calling Syntax

CALL Runs_test (N,Direction,V,P)

Input Parameters

N	number of random deviates used. The value of N should be 4000 or more.
Direction	Direction = 1 means an ascending run. Direction = −1 means a descending run.

Output Parameters

V	Chi-square statistic. Since adjacent runs are not independent, a standard Chi-square test cannot be used here. A special test, with six degrees of freedom is used instead.
P	Right-tailed probability; Prob (X>V).

Algorithm

In this algorithm, we examine the length of monotone subsequences of an original sequence of random numbers; that is, segments which are increasing or decreasing.

1. Calculate the increasing (or decreasing) run lengths and count how many runs have length 1, 2, ..., 6 or greater.
2. Since adjacent runs are not independent, we cannot apply a standard Chi-square test to the above data. Instead, we calculate a special statistic V (see Ref. 1, p. 61) which should have the Chi-square distribution with six degrees of freedom, when N is large. The value of N should be at least 4000 for a valid test. This test may also be used for decreasing runs.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 60-61.

(TSERAL)
Serial Test

Description

This subprogram tests whether pairs of successive numbers are uniformly distributed in an independent manner.

File Name
"TSERAL"

Calling Syntax
CALL Serial_test (N,D,D_squared,V,P)

Input Parameters

N	number of uniform random numbers to be tested.
D	number of digits permitted; 5 or 10 is a reasonable number here.
D_squared	D*D; this must be passed as a parameter to allow for dynamic allocation of arrays.

Output Parameters

V	Chi-square statistic. V is expected to have the Chi-square distribution with (D * D − 1) degrees of freedom.
P	right-tailed probability; Prob(X>V).

Algorithm

Given n = total number of uniform random numbers.

d = number of digits permitted; that is, the deviates created are used to create integers 1,2..., d.

yj = jth random integer.

Then for each pair of integers (q,r) with 0< = q, r<d, count the number of times the pair

(y2j,y2j + 1) = (q,r) occurs, for 0< = j<n.

Finally, apply the Chi-square test to these k = d*d equi-probable categories with probability 1/(d*d) in each case.

Special Considerations

1. The number of digits permitted may be chosen as any convenient number. But care must be taken since a valid Chi-square test should have n large compared to k; that is, n>5*d*d at least.

So, if

d = 10 then n>500
d = 20 then n>2000

etc.

2. This test may easily be adapted to triples, quadruples, etc., instead of pairs. But the value of d must be severely limited in order to avoid having too many categories. Frequently, in this case, less exact tests, such as the poker test or the maximum t test are used instead.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 55-66.

(SPCTRL)
Spectral Test

Description

This test is used in theoretically determining the value of coefficient A, given the word size of the computer, M, in the linear congruential model described in the General Information section of this manual. The value of A is crucial in setting up a good uniform random number generator. This is by far the most powerful test currently available on any sized machine. It tends to measure the statistical independence of adjacent n-tuples of numbers and is generally applied for N = 2,3,4 and perhaps a few higher values of N.

File Name
"SPCTRL"

Calling Syntax
CALL Spectral (A,M,N,Info,Q,V,Cn)

Input Parameters

A — the multiplier to be tested. It is essential that the linear congruential sequence be of maximal period.

M — modulus used in the model; in our case, M≤2^49− 1.

N — size of n-tuple to be measured. This test is generally applied for N = 2,3, 4 and perhaps a few higher values of N.

Info — intermediate information on program execution each time a particular section of code has been entered as well as total number of iterations required for convergence can be printed out at the user's option:

Info = 1 = < print out intermediate information.
Info = 0 = > do not print out the information.

Output Parameters

Q — V ↑ 2, equals the wave number squared.

V — smallest non-zero wave number in the spectrum.

$$Cn = \frac{PI \uparrow (N/2) * V \uparrow N}{(N/2)! * M}$$

Special Considerations

1. Since BASIC string routines are used to perform the multi-precision arithmetic, this program is very slow.
2. The subprogram allows at most 12 digits for A and M. If larger numbers are desired, some parameters must be changed to strings before entering the routine.

Change:

```
SUB Spectral (A,M,N,Info,Q,V,Cn)
DIM - - - - - - -
Coef$ = VAL$(A)
CALL Clean-up (Coef$)
Base$ = VAL$(M)
CALL Clean-up (Base$)
.
.
.
```

To:

```
SUB Spectral (Coef$,Base$,N,Info,Q,V,Cn)
```

3. As suggested in the literature, the driver has been set up for N = 2,3,4,5,6.
4. The multi-precision arithmetic routines are set up as independent subprograms so that the user may apply them to other contexts as well. Presently, each of these routines allows for up to 90 digits of accuracy. This can be increased simply by changing the DIM statements at the beginning of each routine.

Note

This test is quite slow. It is not unusual for it to run for a couple of hours with one pair.

5. The program has been set up with n-tuples of size 2, 3, 4, 5 and 6. For each of these values, the quantity Cn is calculated. Large values of Cn correspond to randomness, small values correspond to nonrandomness. Knuth suggests that the multiplier A passes the spectral test if the Cn values are all greater than or equal to 0.1, and it passes the test with flying colors if all are greater than or equal to 1.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Vol. II, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 69-100.

Elementary Sampling Techniques

Object of Programs

This section provides some elementary sampling and shuffling techniques. Independent subprograms with optional driver routines are provided.

(SSEL)
Selection Sampling

Description

Given a set of N objects, this program will select n of them at random in an unbiased manner (a simple random sample without replacement).

File Name
"SSEL"

Calling Syntax
CALL Sel_sampling (T_number,S_number,X(*))

Input Parameters

T_number	total number of records in the set.
S_number	number of records to be selected.

Output Parameters

X(*)	array of size (1:N) containing the index numbers of the records to be sampled.

Algorithm

To select n records at random from a set of N, where 0<n< = N:

1. Set t = 0, m = 0.
2. Generate a random number U, uniformly distributed between zero and one.
3. If (N – t)*U> = (n – m), then go to step 5.
   Else go to step 4.
4. Select the next record index for the sample.
   m = m + 1.
   t = t + 1.
   If m<n then go to step 2.
   Else the sample is complete and the algorithm terminates.
5. Skip the next record index.
   t = t + 1
   Go to step 2.

Special Considerations

1. In order to avoid connections between samples obtained on different runs, care must be taken to use different starting seeds each time this program is run. RND (using RANDOMIZE) allows for this. The seed can either be initialized in the calling program or the subprogram itself.

   A simple way of initializing different seeds for different runs is to do the following: use the digits from the month, day, and time that the program is run as the seed. For example, if you are running the program on June 19 at 9:47 am, then your seed would be 6190947.

Reference

1. Knuth, Donald E., The Art of Computer Programming. Vol. II, Seminumerical Algorithms, Reading, Mass.: Addison-Wesley, 1969, p. 122.

(SSHUFL)
Shuffling

Description

Given an array of numbers, this program randomly shuffles the array.

File Name
"SSHUFL"

Calling Syntax
CALL Sshuffle (N,X(*))

Input Parameters

N	number of digits in the array to be shuffled.
X(*)	array of dimension (1:N) containing the digits to be shuffled.

Output Parameters

X(*)	array of dimension (1:N) containing the shuffled digits.

Algorithm

Let X1,X2, ..., Xt be a set of t numbers to be shuffled.

1. Set: j = t.
2. Generate a random number U, uniformly distributed between zero and one.
3. Set: k = greatest integer in [j*U + 1]. Hence, k is a random integer between i and j. Exchange Xk and Xj.
4. j = j − 1.
   If j>1 then return to step 2.
   Else the algorithm terminates at this point.

Reference

1. Knuth, Donald E., The Art of Computer Programming, Volume 2, Seminumerical Algorithms. Reading, Mass.: Addison-Wesley, 1969, p. 124-125.

Notes

Appendix

Changes Necessary For Larger Data Sets

CAUTION

INCREASING THE SIZE OF THE DATA SET MAY CAUSE A PROBLEM. THERE MAY NOT BE ENOUGH ROOM ON THE PROGRAM DISC TO STORE THE ENLARGED DATA SET. TO FIND OUT, PROCEED AS FOLLOWS.

A. Perform the following check on each of your program tapes or discs (excluding Monte Carlo Random Number Generator):

1. Make sure nothing of value is in the scratch file "DATA". If there is, use the STORE routine to save it.
2. Type: PURGE "DATA"
3. Press: EXECUTE
4. Type: CREATE "DATA", 2 + (8*n) DIV 1280,1280 where n is the maximum number of data values you wish to use in the statistics routines (and is equal to number of variables times number of observations per variable).
5. Press: EXECUTE

In addition, follow the above procedure for the file named "BACKUP" on Basic Statistics and Data Manipulation.

If you obtain an error using the above procedure on any of the program tapes or discs, you must transfer all data to a larger media in order to expand the data set.

B. Make the following change to Basic Statistics and Data Manipulation:

1. Type: . LOAD"FILE1"
2. Press: . EXECUTE
3. Type: EDIT 80
4. Press: EXECUTE
5. By editing, make the line read
   Mno = n
   where n is the maximum number of data values you wish to use in the statistics routines. This must be less than or equal to 1500.
6. Press: ENTER
7. Press: shift RESET
8. Type: PURGE "FILE1"
9. Press: EXECUTE
10. Type: STORE "FILE1"
11. Press: EXECUTE

Note

Maximum number of variables is 50 and cannot be changed by the user.

Statistics Library Data Formats

The following is a description of the data format used in the Statistics Library. Also included is an explanation of the steps you need to perform to have a program create data compatible with the library.

Method 1 Numeric Data Only

If you wish to have another program, write a data file that is compatible with the library. It is important to note that the actual numeric data could be written in one of two forms:

Observations

Variables	0_1	0_2	0_3	0_4	···	0_N
V_1						
V_2						
V_3						
⋮						
V_p						

OR

Variables

Observation	V_1	V_2	V_3	···	V_p
0_1					
0_2					
0_3					
⋮					
0_N					

The statistics library will prompt you for additional information such as sample size (n), number of variables (p), title of the data set, and names of the variables.

The statements needed to store the data are as follows:

```
05 OPTION BASE 1
10 P=3               ! P=no. of variables
20 N=10              ! N=no. of observations
30 ALLOCATE X(P,N)   ! THIS COULD BE X(N,P)
40 !
50 ! Put data into matrix X
60 !
70 CREATE BDAT "FILE ",INT((8*P*N)/1280)+2,1280 ! 8 bytes per entry and
80 ASSIGN @File1 TO "FILE1"                     ! 1280 bytes per logical
90 OUTPUT @File1;X(*)                           ! record
100 ASSIGN @File1 TO *
110 END
```

Method 2 Numeric Data and Descriptive Data

If you wish to have another program, write a data file that is compatible with the library and if you wish to have it store descriptive information as well, you need to prepare the file in a slightly different manner.

The following data is stored in record 1 of the data file:

Data set title	T$[80]	
Number of observations	No	
Number of variables	Nv	(max. is 50)
Variable names	Vn$(50)[10]	
Number of subfiles	Ns	(max. is 20)
Subfile names	Sn$(20)[10]	
Subfile characterizations	Sc(20)	

Note

No, Nv, Ns, and the array Sc(*) should be declared in real precision.

Starting with record 2, the Statistics Library expects to find the data array.

The statements needed to store the data are as follows:

```
05 OPTION BASE 1
10 P=3              ! P=no. of variables
20 N=10             ! N=no. of observations
30 ALLOCATE X(P,N)
35 DIM  T$[80], Vn$(50)[10], Sn$(20)[10], Sc(20)
40 !
50 ! Put data into matrix X and descriptive data into other variables
60 !
70 CREATE BDAT "FILE1",INT((8*P*N)/1280)+2,1280
80 ASSIGN @File1 TO "FILE1"
85 OUTPUT @File,1;T$,No,Nv,Vn$(*),Ns,Sn$(*),Sc(*)  ! Write record 1
90 OUTPUT @File,2;X(*)                              ! Write records 2,3,...
100 ASSIGN @File1 to *
110 END
```

When using this format and the Statistics Library asks you the question, "Was the data stored by the BS&DM system?", answer Yes. This will tell the library to expect the header record as record #1.

Statistical Tables

Quantiles of the Spearman Test Statistic[a]

n	p = .900	.950	.975	.990	.995	.999
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.7818	.8667
11	.4182	.5273	.6091	.7000	.7455	.8364
12	.3986	.4965	.5804	.6713	.7273	.8182
13	.3791	.4780	.5549	.6429	.6978	.7912
14	.3626	.4593	.5341	.6220	.6747	.7670
15	.3500	.4429	.5179	.6000	.6536	.7464
16	.3382	.4265	.5000	.5824	.6324	.7265
17	.3260	.4118	.4853	.5637	.6152	.7083
18	.3148	.3994	.4716	.5480	.5975	.6904
19	.3070	.3895	.4579	.5333	.5825	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

[a] The entries in this table are selected quantiles w_p of the Spearman rank correlation coefficient ρ when used as a test statistic. The lower quantiles may be obtained from the equation

$$w_p = -w_{1-p}$$

The critical region corresponds to values of ρ smaller than (or greater than) but not including the appropriate quantile. Note that the median of ρ is 0.

This table was reprinted from Practical Nonparametric Statistics by W.J. Conover, with permission from John Wiley and Sons, Inc., and authors Dr. Gerald J. Glasser and Dr. Winter.

Quantiles of the Wilcoxon Signed Ranks Test Statistic[a]

	$w_{.005}$	$w_{.01}$	$w_{.025}$	$w_{.05}$	$w_{.10}$	$w_{.20}$	$w_{.30}$	$w_{.40}$	$w_{.50}$	$\frac{n(n+1)}{2}$
$n = 4$	0	0	0	0	1	3	3	4	5	10
5	0	0	0	1	3	4	5	6	7.5	15
6	0	0	1	3	4	6	8	9	10.5	21
7	0	1	3	4	6	9	11	12	14	28
8	1	2	4	6	9	12	14	16	18	36
9	2	4	6	9	11	15	18	20	22.5	45
10	4	6	9	11	15	19	22	25	27.5	55
11	6	8	11	14	18	23	27	30	33	66
12	8	10	14	18	22	28	32	36	39	78
13	10	13	18	22	27	33	38	42	45.5	91
14	13	16	22	26	32	39	44	48	52.5	105
15	16	20	26	31	37	45	51	55	60	120
16	20	24	30	36	43	51	58	63	68	136
17	24	28	35	42	49	58	65	71	76.5	153
18	28	33	41	48	56	66	73	80	85.5	171
19	33	38	47	54	63	74	82	89	95	190
20	38	44	53	61	70	82	91	98	105	210

[a] The entries in this table are quantiles w_p of the Wilcoxon signed ranks test statistic T, for selected values of $p \leq .50$. Quantiles w_p for $p > .50$ may be computed from the equation

$$w_p = n(n + 1)/2 - w_{1-p}$$

where $n(n + 1)/2$ is given in the right hand column in the table. Note that $P(T < w_p) \leq p$ and $P(T > w_p) \leq 1 - p$ if H_0 is true. Critical regions correspond to values of T less than (or greater than) but not including the appropriate quantile.

This table was reprinted from the Journal of the American Statistical Association, Dr. Robert L. McCornack author, and with the permission of the American Statistical Association.

Quantiles of the Kolmogorov Test Statistic[a]

One-Sided Test $p = .90$	.95	.975	.99	.995		$p = .90$	.95	.975	.99	.995
Two-Sided Test $p = .80$	.90	.95	.98	.99		$p = .80$	.90	.95	.98	.99
$n = 1$ .900	.950	.975	.990	.995	$n = 21$	.226	.259	.287	.321	.344
2 .684	.776	.842	.900	.929	22	.221	.253	.281	.314	.337
3 .565	.636	.708	.785	.829	23	.216	.247	.275	.307	.330
4 .493	.565	.624	.689	.734	24	.212	.242	.269	.301	.323
5 .447	.509	.563	.627	.669	25	.208	.238	.264	.295	.317
6 .410	.468	.519	.577	.617	26	.204	.233	.259	.290	.311
7 .381	.436	.483	.538	.576	27	.200	.229	.254	.284	.305
8 .358	.410	.454	.507	.542	28	.197	.225	.250	.279	.300
9 .339	.387	.430	.480	.513	29	.193	.221	.246	.275	.295
10 .323	.369	.409	.457	.489	30	.190	.218	.242	.270	.290
11 .308	.352	.391	.437	.468	31	.187	.214	.238	.266	.285
12 .296	.338	.375	.419	.449	32	.184	.211	.234	.262	.281
13 .285	.325	.361	.404	.432	33	.182	.208	.231	.258	.277
14 .275	.314	.349	.390	.418	34	.179	.205	.227	.254	.273
15 .266	.304	.338	.377	.404	35	.177	.202	.224	.251	.269
16 .258	.295	.327	.366	.392	36	.174	.199	.221	.247	.265
17 .250	.286	.318	.355	.381	37	.172	.196	.218	.244	.262
18 .244	.279	.309	.346	.371	38	.170	.194	.215	.241	.258
19 .237	.271	.301	.337	.361	39	.168	.191	.213	.238	.255
20 .232	.265	.294	.329	.352	40	.165	.189	.210	.235	.252
			Approximation for $n > 40$			$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

[a] The entries in this table are selected quantiles w_p of the Kolmogorov test statistics T_1, T_1^+, and T_1^- as defined by (6.1.1) for two-sided tests and by (6.1.2) and (6.1.3) for one-sided tests. Reject H_0 at the level α if T exceeds the $1 - \alpha$ quantile given in this table. These quantiles are exact for $n \leq 20$ in the two-tailed test. The other quantiles are approximations which are equal to the exact quantiles in most cases.

This table was reprinted from the Journal of the American Statistical Association with the permission of the American Statistical Association, author Dr. J.L. Miller.

Quantiles of the Mann-Whitney Test Statistic

n	*p*	*m* = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	.005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
2	.01	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	2	2
	.025	0	0	0	0	0	0	1	1	1	1	2	2	2	2	2	3	3	3	3
	.05	0	0	0	1	1	1	2	2	2	2	3	3	4	4	4	4	5	5	5
	.10	0	1	1	2	2	2	3	3	4	4	5	5	5	6	6	7	7	8	8
	.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
	.005	0	0	0	0	0	0	0	1	1	1	2	2	2	3	3	3	3	4	4
3	.01	0	0	0	0	0	1	1	2	2	2	3	3	3	4	4	5	5	5	6
	.025	0	0	0	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9
	.05	0	1	1	2	3	3	4	5	5	6	6	7	8	8	9	10	10	11	12
	.10	1	2	2	3	4	5	6	6	7	8	9	10	11	11	12	13	14	15	16
	.001	0	0	0	0	0	0	0	0	1	1	1	2	2	2	3	3	4	4	4
	.005	0	0	0	0	1	1	2	2	3	3	4	4	5	6	6	7	7	8	9
4	.01	0	0	0	1	2	2	3	4	4	5	6	6	7	9	8	9	10	10	11
	.025	0	0	1	2	3	4	5	5	6	7	8	9	10	11	12	12	13	14	15
	.05	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19
	.10	1	2	4	5	6	7	8	10	11	12	13	14	16	17	18	19	21	22	23
	.001	0	0	0	0	0	0	1	2	2	3	3	4	4	5	6	6	7	8	8
	.005	0	0	0	1	2	2	3	4	5	6	7	8	8	9	10	11	12	13	14
5	.01	0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	.025	0	1	2	3	4	6	7	8	9	10	12	13	14	15	16	18	19	20	21
	.05	1	2	3	5	6	7	9	10	12	13	14	16	17	19	20	21	23	24	26
	.10	2	3	5	6	8	9	11	13	14	16	18	19	21	23	24	26	28	29	31
	.001	0	0	0	0	0	0	2	3	4	5	5	6	7	8	9	10	11	12	13
	.005	0	0	1	2	3	4	5	6	7	8	10	11	12	13	14	16	17	18	19
6	.01	0	0	2	3	4	5	7	8	9	10	12	13	14	16	17	19	20	21	23
	.025	0	2	3	4	6	7	9	11	12	14	15	17	18	20	22	23	25	26	28
	.05	1	3	4	6	8	9	11	13	15	17	18	20	22	24	26	27	29	31	33
	.10	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	35	37	39
	.001	0	0	0	0	1	2	3	4	6	7	8	9	10	11	12	14	15	16	17
	.005	0	0	1	2	4	5	7	8	10	11	13	14	16	17	19	20	22	23	25
7	.01	0	1	2	4	5	7	8	10	12	13	15	17	18	20	22	24	25	27	29
	.025	0	2	4	6	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35
	.05	1	3	5	7	9	12	14	16	18	20	22	25	27	29	31	34	36	38	40
	.10	2	5	7	9	12	14	17	19	22	24	27	29	32	34	37	39	42	44	47
	.001	0	0	0	1	2	3	5	6	7	9	10	12	13	15	16	18	19	21	22
	.005	0	0	2	3	5	7	8	10	12	14	16	18	19	21	23	25	27	29	31
8	.01	0	1	3	5	7	8	10	12	14	16	18	21	23	25	27	29	31	33	35
	.025	1	3	5	7	9	11	14	16	18	20	23	25	27	30	32	35	37	39	42
	.05	2	4	6	9	11	14	16	19	21	24	27	29	32	34	37	40	42	45	48
	.10	3	6	8	11	14	17	20	23	25	28	31	34	37	40	43	46	49	52	55
	.001	0	0	0	2	3	4	6	8	9	11	13	15	16	18	20	22	24	26	27
	.005	0	1	2	4	6	8	10	12	14	17	19	21	23	25	28	30	32	34	37
9	.01	0	2	4	6	8	10	12	15	17	19	22	24	27	29	32	34	37	39	41
	.025	1	3	5	8	11	13	16	18	21	24	27	29	32	35	38	40	43	46	49
	.05	2	5	7	10	13	16	19	22	25	28	31	34	37	40	43	46	49	52	55
	.10	3	6	10	13	16	19	23	26	29	32	36	39	42	46	49	53	56	59	63
	.001	0	0	1	2	4	6	7	9	11	13	15	18	20	22	24	26	28	30	33
	.005	0	1	3	5	7	10	12	14	17	19	22	25	27	30	32	35	38	40	43
10	.01	0	2	4	7	9	12	14	17	20	23	25	28	31	34	37	39	42	45	48
	.025	1	4	6	9	12	15	18	21	24	27	30	34	37	40	43	46	49	53	56
	.05	2	5	8	12	15	18	21	25	28	32	35	38	42	45	49	52	56	59	63
	.10	4	7	11	14	18	22	25	29	33	37	40	44	48	52	55	59	63	67	71
	.001	0	0	1	3	5	7	9	11	13	16	18	21	23	25	28	30	33	35	38
	.005	0	1	3	6	8	11	14	17	19	22	25	28	31	34	37	40	43	46	49
11	.01	0	2	5	8	10	13	16	19	23	26	29	32	35	38	42	45	48	51	54
	.025	1	4	7	10	14	17	20	24	27	31	34	38	41	45	48	52	56	59	63
	.05	2	6	9	13	17	20	24	28	32	35	39	43	47	51	55	58	62	66	70
	.10	4	8	12	16	20	24	28	32	37	41	45	49	53	58	62	66	70	74	79

Quantiles of the Mann-Whitney Test Statistic (continued)

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	.001	0	0	1	3	5	8	10	13	15	18	21	24	26	29	32	35	38	41	43
	.005	0	2	4	7	10	13	16	19	22	25	28	32	35	38	42	45	48	52	55
12	.01	0	3	6	9	12	15	18	22	25	29	32	36	39	43	47	50	54	57	61
	.025	2	5	8	12	15	19	23	27	30	34	38	42	46	50	54	58	62	66	70
	.05	3	6	10	14	18	22	27	31	35	39	43	48	52	56	61	65	69	73	78
	.10	5	9	13	18	22	27	31	36	40	45	50	54	59	64	68	73	78	82	87
	.001	0	0	2	4	6	9	12	15	18	21	24	27	30	33	36	39	43	46	49
	.005	0	2	4	8	11	14	18	21	25	28	32	35	39	43	46	50	54	58	61
13	.01	1	3	6	10	13	17	21	24	28	32	36	40	44	48	52	56	60	64	68
	.025	2	5	9	13	17	21	25	29	34	38	42	46	51	55	60	64	68	73	77
	.05	3	7	11	16	20	25	29	34	38	43	48	52	57	62	66	71	76	81	85
	.10	5	10	14	19	24	29	34	39	44	49	54	59	64	69	75	80	85	90	95
	.001	0	0	2	4	7	10	13	16	20	23	26	30	33	37	40	44	47	51	55
	.005	0	2	5	8	12	16	19	23	27	31	35	39	43	47	51	55	59	64	68
14	.01	1	3	7	11	14	18	23	27	31	35	39	44	48	52	57	61	66	70	74
	.025	2	6	10	14	18	23	27	32	37	41	46	51	56	60	65	70	75	79	84
	.05	4	8	12	17	22	27	32	37	42	47	52	57	62	67	72	78	83	88	93
	.10	5	11	16	21	26	32	37	42	48	53	59	64	70	75	81	86	92	98	103
	.001	0	0	2	5	8	11	15	18	22	25	29	33	37	41	44	48	52	56	60
	.005	0	3	6	9	13	17	21	25	30	34	38	43	47	52	56	61	65	70	74
15	.01	1	4	8	12	16	20	25	29	34	38	43	48	52	57	62	67	71	76	81
	.025	2	6	11	15	20	25	30	35	40	45	50	55	60	65	71	76	81	86	91
	.05	4	8	13	19	24	29	34	40	45	51	56	62	67	73	78	84	89	95	101
	.10	6	11	17	23	28	34	40	46	52	58	64	69	75	81	87	93	99	105	111
	.001	0	0	3	6	9	12	16	20	24	28	32	36	40	44	49	53	57	61	66
	.005	0	3	6	10	14	19	23	28	32	37	42	46	51	56	61	66	71	75	80
16	.01	1	4	8	13	17	22	27	32	37	42	47	52	57	62	67	72	77	83	88
	.025	2	7	12	16	22	27	32	38	43	48	54	60	65	71	76	82	87	93	99
	.05	4	9	15	20	26	31	37	43	49	55	61	66	72	78	84	90	96	102	108
	.10	6	12	18	24	30	37	43	49	55	62	68	75	81	87	94	100	107	113	120
	.001	0	1	3	6	10	14	18	22	26	30	35	39	44	48	53	58	62	67	71
	.005	0	3	7	11	16	20	25	30	35	40	45	50	55	61	66	71	76	82	87
17	.01	1	5	9	14	19	24	29	34	39	45	50	56	61	67	72	78	83	89	94
	.025	3	7	12	18	23	29	35	40	46	52	58	64	70	76	82	88	94	100	106
	.05	4	10	16	21	27	34	40	46	52	58	65	71	78	84	90	97	103	110	116
	.10	7	13	19	26	32	39	46	53	59	66	73	80	86	93	100	107	114	121	128
	.001	0	1	4	7	11	15	19	24	28	33	38	43	47	52	57	62	67	72	77
	.005	0	3	7	12	17	22	27	32	38	43	48	54	59	65	71	76	82	88	93
18	.01	1	5	10	15	20	25	31	37	42	48	54	60	66	71	77	83	89	95	101
	.025	3	8	13	19	25	31	37	43	49	56	62	68	75	81	87	94	100	107	113
	.05	5	10	17	23	29	36	42	49	56	62	69	76	83	89	96	103	110	117	124
	.10	7	14	21	28	35	42	49	56	63	70	78	85	92	99	107	114	121	129	136
	.001	0	1	4	8	12	16	21	26	30	35	41	46	51	56	61	67	72	78	83
	.005	1	4	8	13	18	23	29	34	40	46	52	58	64	70	75	82	88	94	100
19	.01	2	5	10	16	21	27	33	39	45	51	57	64	70	76	83	89	95	102	108
	.025	3	8	14	20	26	33	39	46	53	59	66	73	79	86	93	100	107	114	120
	.05	5	11	18	24	31	38	45	52	59	66	73	81	88	95	102	110	117	124	131
	.10	8	15	22	29	37	44	52	59	67	74	82	90	98	105	113	121	129	136	144
	.001	0	1	4	8	13	17	22	27	33	38	43	49	55	60	66	71	77	83	89
	.005	1	4	9	14	19	25	31	37	43	49	55	61	68	74	80	87	93	100	106
20	.01	2	6	11	17	23	29	35	41	48	54	61	68	74	81	88	94	101	108	115
	.025	3	9	15	21	28	35	42	49	56	63	70	77	84	91	99	106	113	120	128
	.05	5	12	19	26	33	40	48	55	63	70	78	85	93	101	108	116	124	131	139
	.10	8	16	23	31	39	47	55	63	71	79	87	95	103	111	120	128	136	144	152

This table was reprinted from Practical Nonparametric Statistics by W.J. Conover, with permission from John Wiley and Sons, Inc., and author L.R. Verdooren.

Percentage Points of the Duncan New Multiple Range Test

n_2 \ p	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47	3.47	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47	3.47	3.47
24	2.02	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47	3.47	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47	3.47	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47	3.53	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47	3.61	3.67

*Using special protection levels based on degrees of freedom.

This table was reprinted from Biometrics, Vol. II with the permission of the Biometric Society and author D.B. Duncan.

Percentage Points of the Studentized Range, $q=(x_n-x_1)/s_\nu$.

Upper 10% points

ν \ n	2	3	4	5	6	7	8	9	10
1	8·93	13·44	16·36	18·49	20·15	21·51	22·64	23·62	24·48
2	4·13	5·73	6·77	7·54	8·14	8·63	9·05	9·41	9·72
3	3·33	4·47	5·20	5·74	6·16	6·51	6·81	7·06	7·29
4	3·01	3·98	4·59	5·03	5·39	5·68	5·93	6·14	6·33
5	2·85	3·72	4·26	4·66	4·98	5·24	5·46	5·65	5·82
6	2·75	3·56	4·07	4·44	4·73	4·97	5·17	5·34	5·50
7	2·68	3·45	3·93	4·28	4·55	4·78	4·97	5·14	5·28
8	2·63	3·37	3·83	4·17	4·43	4·65	4·83	4·99	5·13
9	2·59	3·32	3·76	4·08	4·34	4·54	4·72	4·87	5·01
10	2·56	3·27	3·70	4·02	4·26	4·47	4·64	4·78	4·91
11	2·54	3·23	3·66	3·96	4·20	4·40	4·57	4·71	4·84
12	2·52	3·20	3·62	3·92	4·16	4·35	4·51	4·65	4·78
13	2·50	3·18	3·59	3·88	4·12	4·30	4·46	4·60	4·72
14	2·49	3·16	3·56	3·85	4·08	4·27	4·42	4·56	4·68
15	2·48	3·14	3·54	3·83	4·05	4·23	4·39	4·52	4·64
16	2·47	3·12	3·52	3·80	4·03	4·21	4·36	4·49	4·61
17	2·46	3·11	3·50	3·78	4·00	4·18	4·33	4·46	4·58
18	2·45	3·10	3·49	3·77	3·98	4·16	4·31	4·44	4·55
19	2·45	3·09	3·47	3·75	3·97	4·14	4·29	4·42	4·53
20	2·44	3·08	3·46	3·74	3·95	4·12	4·27	4·40	4·51
24	2·42	3·05	3·42	3·69	3·90	4·07	4·21	4·34	4·44
30	2·40	3·02	3·39	3·65	3·85	4·02	4·16	4·28	4·38
40	2·38	2·99	3·35	3·60	3·80	3·96	4·10	4·21	4·32
60	2·36	2·96	3·31	3·56	3·75	3·91	4·04	4·16	4·25
120	2·34	2·93	3·28	3·52	3·71	3·86	3·99	4·10	4·19
∞	2·33	2·90	3·24	3·48	3·66	3·81	3·93	4·04	4·13

ν \ n	11	12	13	14	15	16	17	18	19	20
1	25·24	25·92	26·54	27·10	27·62	28·10	28·54	28·96	29·35	29·71
2	10·01	10·26	10·49	10·70	10·89	11·07	11·24	11·39	11·54	11·68
3	7·49	7·67	7·83	7·98	8·12	8·25	8·37	8·48	8·58	8·68
4	6·49	6·65	6·78	6·91	7·02	7·13	7·23	7·33	7·41	7·50
5	5·97	6·10	6·22	6·34	6·44	6·54	6·63	6·71	6·79	6·86
6	5·64	5·76	5·87	5·98	6·07	6·16	6·25	6·32	6·40	6·47
7	5·41	5·53	5·64	5·74	5·83	5·91	5·99	6·06	6·13	6·19
8	5·25	5·36	5·46	5·56	5·64	5·72	5·80	5·87	5·93	6·00
9	5·13	5·23	5·33	5·42	5·51	5·58	5·66	5·72	5·79	5·85
10	5·03	5·13	5·23	5·32	5·40	5·47	5·54	5·61	5·67	5·73
11	4·95	5·05	5·15	5·23	5·31	5·38	5·45	5·51	5·57	5·63
12	4·89	4·99	5·08	5·16	5·24	5·31	5·37	5·44	5·49	5·55
13	4·83	4·93	5·02	5·10	5·18	5·25	5·31	5·37	5·43	5·48
14	4·79	4·88	4·97	5·05	5·12	5·19	5·26	5·32	5·37	5·43
15	4·75	4·84	4·93	5·01	5·08	5·15	5·21	5·27	5·32	5·38
16	4·71	4·81	4·89	4·97	5·04	5·11	5·17	5·23	5·28	5·33
17	4·68	4·77	4·86	4·93	5·01	5·07	5·13	5·19	5·24	5·30
18	4·65	4·75	4·83	4·90	4·98	5·04	5·10	5·16	5·21	5·26
19	4·63	4·72	4·80	4·88	4·95	5·01	5·07	5·13	5·18	5·23
20	4·61	4·70	4·78	4·85	4·92	4·99	5·05	5·10	5·16	5·20
24	4·54	4·63	4·71	4·78	4·85	4·91	4·97	5·02	5·07	5·12
30	4·47	4·56	4·64	4·71	4·77	4·83	4·89	4·94	4·99	5·03
40	4·41	4·49	4·56	4·63	4·69	4·75	4·81	4·86	4·90	4·95
60	4·34	4·42	4·49	4·56	4·62	4·67	4·73	4·78	4·82	4·86
120	4·28	4·35	4·42	4·48	4·54	4·60	4·65	4·69	4·74	4·78
∞	4·21	4·28	4·35	4·41	4·47	4·52	4·57	4·61	4·65	4·69

n: size of sample from which range obtained. ν: degrees of freedom of independent s_ν.

Percentage Points of the Studentized Range, $q=(x_n-x_1)/s_\nu$. (continued)

Upper 5 % points

ν \ n	2	3	4	5	6	7	8	9	10
1	17·97	26·98	32·82	37·08	40·41	43·12	45·40	47·36	49·07
2	6·08	8·33	9·80	10·88	11·74	12·44	13·03	13·54	13·99
3	4·50	5·91	6·82	7·50	8·04	8·48	8·85	9·18	9·46
4	3·93	5·04	5·76	6·29	6·71	7·05	7·35	7·60	7·83
5	3·64	4·60	5·22	5·67	6·03	6·33	6·58	6·80	6·99
6	3·46	4·34	4·90	5·30	5·63	5·90	6·12	6·32	6·49
7	3·34	4·16	4·68	5·06	5·36	5·61	5·82	6·00	6·16
8	3·26	4·04	4·53	4·89	5·17	5·40	5·60	5·77	5·92
9	3·20	3·95	4·41	4·76	5·02	5·24	5·43	5·59	5·74
10	3·15	3·88	4·33	4·65	4·91	5·12	5·30	5·46	5·60
11	3·11	3·82	4·26	4·57	4·82	5·03	5·20	5·35	5·49
12	3·08	3·77	4·20	4·51	4·75	4·95	5·12	5·27	5·39
13	3·06	3·73	4·15	4·45	4·69	4·88	5·05	5·19	5·32
14	3·03	3·70	4·11	4·41	4·64	4·83	4·99	5·13	5·25
15	3·01	3·67	4·08	4·37	4·59	4·78	4·94	5·08	5·20
16	3·00	3·65	4·05	4·33	4·56	4·74	4·90	5·03	5·15
17	2·98	3·63	4·02	4·30	4·52	4·70	4·86	4·99	5·11
18	2·97	3·61	4·00	4·28	4·49	4·67	4·82	4·96	5·07
19	2·96	3·59	3·98	4·25	4·47	4·65	4·79	4·92	5·04
20	2·95	3·58	3·96	4·23	4·45	4·62	4·77	4·90	5·01
24	2·92	3·53	3·90	4·17	4·37	4·54	4·68	4·81	4·92
30	2·89	3·49	3·85	4·10	4·30	4·46	4·60	4·72	4·82
40	2·86	3·44	3·79	4·04	4·23	4·39	4·52	4·63	4·73
60	2·83	3·40	3·74	3·98	4·16	4·31	4·44	4·55	4·65
120	2·80	3·36	3·68	3·92	4·10	4·24	4·36	4·47	4·56
∞	2·77	3·31	3·63	3·86	4·03	4·17	4·29	4·39	4·47

ν \ n	11	12	13	14	15	16	17	18	19	20
1	50·59	51·96	53·20	54·33	55·36	56·32	57·22	58·04	58·83	59·56
2	14·39	14·75	15·08	15·38	15·65	15·91	16·14	16·37	16·57	16·77
3	9·72	9·95	10·15	10·35	10·52	10·69	10·84	10·98	11·11	11·24
4	8·03	8·21	8·37	8·52	8·66	8·79	8·91	9·03	9·13	9·23
5	7·17	7·32	7·47	7·60	7·72	7·83	7·93	8·03	8·12	8·21
6	6·65	6·79	6·92	7·03	7·14	7·24	7·34	7·43	7·51	7·59
7	6·30	6·43	6·55	6·66	6·76	6·85	6·94	7·02	7·10	7·17
8	6·05	6·18	6·29	6·39	6·48	6·57	6·65	6·73	6·80	6·87
9	5·87	5·98	6·09	6·19	6·28	6·36	6·44	6·51	6·58	6·64
10	5·72	5·83	5·93	6·03	6·11	6·19	6·27	6·34	6·40	6·47
11	5·61	5·71	5·81	5·90	5·98	6·06	6·13	6·20	6·27	6·33
12	5·51	5·61	5·71	5·80	5·88	5·95	6·02	6·09	6·15	6·21
13	5·43	5·53	5·63	5·71	5·79	5·86	5·93	5·99	6·05	6·11
14	5·36	5·46	5·55	5·64	5·71	5·79	5·85	5·91	5·97	6·03
15	5·31	5·40	5·49	5·57	5·65	5·72	5·78	5·85	5·90	5·96
16	5·26	5·35	5·44	5·52	5·59	5·66	5·73	5·79	5·84	5·90
17	5·21	5·31	5·39	5·47	5·54	5·61	5·67	5·73	5·79	5·84
18	5·17	5·27	5·35	5·43	5·50	5·57	5·63	5·69	5·74	5·79
19	5·14	5·23	5·31	5·39	5·46	5·53	5·59	5·65	5·70	5·75
20	5·11	5·20	5·28	5·36	5·43	5·49	5·55	5·61	5·66	5·71
24	5·01	5·10	5·18	5·25	5·32	5·38	5·44	5·49	5·55	5·59
30	4·92	5·00	5·08	5·15	5·21	5·27	5·33	5·38	5·43	5·47
40	4·82	4·90	4·98	5·04	5·11	5·16	5·22	5·27	5·31	5·36
60	4·73	4·81	4·88	4·94	5·00	5·06	5·11	5·15	5·20	5·24
120	4·64	4·71	4·78	4·84	4·90	4·95	5·00	5·04	5·09	5·13
∞	4·55	4·62	4·68	4·74	4·80	4·85	4·89	4·93	4·97	5·01

n: size of sample from which range obtained. ν: degrees of freedom of independent s_ν.

Percentage Points of the Studentized Range, $q=(x_n-x_1)/s_\nu$. (continued)

Upper 1% points

ν \ n	2	3	4	5	6	7	8	9	10
1	90·03	135·0	164·3	185·6	202·2	215·8	227·2	237·0	245·6
2	14·04	19·02	22·29	24·72	26·63	28·20	29·53	30·68	31·69
3	8·26	10·62	12·17	13·33	14·24	15·00	15·64	16·20	16·69
4	6·51	8·12	9·17	9·96	10·58	11·10	11·55	11·93	12·27
5	5·70	6·98	7·80	8·42	8·91	9·32	9·67	9·97	10·24
6	5·24	6·33	7·03	7·56	7·97	8·32	8·61	8·87	9·10
7	4·95	5·92	6·54	7·01	7·37	7·68	7·94	8·17	8·37
8	4·75	5·64	6·20	6·62	6·96	7·24	7·47	7·68	7·86
9	4·60	5·43	5·96	6·35	6·66	6·91	7·13	7·33	7·49
10	4·48	5·27	5·77	6·14	6·43	6·67	6·87	7·05	7·21
11	4·39	5·15	5·62	5·97	6·25	6·48	6·67	6·84	6·99
12	4·32	5·05	5·50	5·84	6·10	6·32	6·51	6·67	6·81
13	4·26	4·96	5·40	5·73	5·98	6·19	6·37	6·53	6·67
14	4·21	4·89	5·32	5·63	5·88	6·08	6·26	6·41	6·54
15	4·17	4·84	5·25	5·56	5·80	5·99	6·16	6·31	6·44
16	4·13	4·79	5·19	5·49	5·72	5·92	6·08	6·22	6·35
17	4·10	4·74	5·14	5·43	5·66	5·85	6·01	6·15	6·27
18	4·07	4·70	5·09	5·38	5·60	5·79	5·94	6·08	6·20
19	4·05	4·67	5·05	5·33	5·55	5·73	5·89	6·02	6·14
20	4·02	4·64	5·02	5·29	5·51	5·69	5·84	5·97	6·09
24	3·96	4·55	4·91	5·17	5·37	5·54	5·69	5·81	5·92
30	3·89	4·45	4·80	5·05	5·24	5·40	5·54	5·65	5·76
40	3·82	4·37	4·70	4·93	5·11	5·26	5·39	5·50	5·60
60	3·76	4·28	4·59	4·82	4·99	5·13	5·25	5·36	5·45
120	3·70	4·20	4·50	4·71	4·87	5·01	5·12	5·21	5·30
∞	3·64	4·12	4·40	4·60	4·76	4·88	4·99	5·08	5·16

ν \ n	11	12	13	14	15	16	17	18	19	20
1	253·2	260·0	266·2	271·8	277·0	281·8	286·3	290·4	294·3	298·0
2	32·59	33·40	34·13	34·81	35·43	36·00	36·53	37·03	37·50	37·95
3	17·13	17·53	17·89	18·22	18·52	18·81	19·07	19·32	19·55	19·77
4	12·57	12·84	13·09	13·32	13·53	13·73	13·91	14·08	14·24	14·40
5	10·48	10·70	10·89	11·08	11·24	11·40	11·55	11·68	11·81	11·93
6	9·30	9·48	9·65	9·81	9·95	10·08	10·21	10·32	10·43	10·54
7	8·55	8·71	8·86	9·00	9·12	9·24	9·35	9·46	9·55	9·65
8	8·03	8·18	8·31	8·44	8·55	8·66	8·76	8·85	8·94	9·03
9	7·65	7·78	7·91	8·03	8·13	8·23	8·33	8·41	8·49	8·57
10	7·36	7·49	7·60	7·71	7·81	7·91	7·99	8·08	8·15	8·23
11	7·13	7·25	7·36	7·46	7·56	7·65	7·73	7·81	7·88	7·95
12	6·94	7·06	7·17	7·26	7·36	7·44	7·52	7·59	7·66	7·73
13	6·79	6·90	7·01	7·10	7·19	7·27	7·35	7·42	7·48	7·55
14	6·66	6·77	6·87	6·96	7·05	7·13	7·20	7·27	7·33	7·39
15	6·55	6·66	6·76	6·84	6·93	7·00	7·07	7·14	7·20	7·26
16	6·46	6·56	6·66	6·74	6·82	6·90	6·97	7·03	7·09	7·15
17	6·38	6·48	6·57	6·66	6·73	6·81	6·87	6·94	7·00	7·05
18	6·31	6·41	6·50	6·58	6·65	6·73	6·79	6·85	6·91	6·97
19	6·25	6·34	6·43	6·51	6·58	6·65	6·72	6·78	6·84	6·89
20	6·19	6·28	6·37	6·45	6·52	6·59	6·65	6·71	6·77	6·82
24	6·02	6·11	6·19	6·26	6·33	6·39	6·45	6·51	6·56	6·61
30	5·85	5·93	6·01	6·08	6·14	6·20	6·26	6·31	6·36	6·41
40	5·69	5·76	5·83	5·90	5·96	6·02	6·07	6·12	6·16	6·21
60	5·53	5·60	5·67	5·73	5·78	5·84	5·89	5·93	5·97	6·01
120	5·37	5·44	5·50	5·56	5·61	5·66	5·71	5·75	5·79	5·83
∞	5·23	5·29	5·35	5·40	5·45	5·49	5·54	5·57	5·61	5·65

This table was reprinted from Biometrika Tables for Statisticians, Vol. 1, 3rd Edition, Table 29, with the permission of the Biometrika Trustees.

The Normal Probability Function

The integral $P(X)$ and ordinate $Z(X)$ in terms of the standardized deviate X

X	$P(X)$	δ +	δ^2 −	$Z(X)$	δ −	δ^2 −
·00	·5000000	39894	0	·3989423	199	399
·01	·5039894	39890	4	·3989223	598	399
·02	·5079783	39882	8	·3988625	997	399
·03	·5119665	39870	12	·3987628	1395	398
·04	·5159534	39854	16	·3986233	1793	398
·05	·5199388	39834	20	·3984439	2191	397
·06	·5239222	39810	24	·3982248	2588	397
·07	·5279032	39782	28	·3979661	2984	396
·08	·5318814	39750	32	·3976677	3379	395
·09	·5358564	39714	36	·3973298	3773	394
·10	·5398278	39675	40	·3969525	4166	393
·11	·5437953	39631	44	·3965360	4558	392
·12	·5477584	39584	48	·3960802	4948	390
·13	·5517168	39532	51	·3955854	5337	389
·14	·5556700	39477	55	·3950517	5724	387
·15	·5596177	39418	59	·3944793	6110	386
·16	·5635595	39355	63	·3938684	6493	384
·17	·5674949	39288	67	·3932190	6875	382
·18	·5714237	39217	71	·3925315	7255	380
·19	·5753454	39143	74	·3918060	7633	378
·20	·5792597	39065	78	·3910427	8008	375
·21	·5831662	38983	82	·3902419	8381	373
·22	·5870644	38897	86	·3894038	8752	371
·23	·5909541	38808	89	·3885286	9120	368
·24	·5948349	38715	93	·3876166	9485	365
·25	·5987063	38618	97	·3866681	9847	362
·26	·6025681	38518	100	·3856834	10207	360
·27	·6064199	38414	104	·3846627	10564	357
·28	·6102612	38306	107	·3836063	10917	354
·29	·6140919	38195	111	·3825146	11268	350
·30	·6179114	38081	114	·3813878	11615	347
·31	·6217195	37963	118	·3802264	11958	344
·32	·6255158	37842	121	·3790305	12298	340
·33	·6293000	37717	125	·3778007	12635	337
·34	·6330717	37589	128	·3765372	12968	333
·35	·6368307	37458	131	·3752403	13297	329
·36	·6405764	37323	135	·3739106	13623	325
·37	·6443088	37185	138	·3725483	13944	322
·38	·6480273	37044	141	·3711539	14262	318
·39	·6517317	36900	144	·3697277	14575	313
·40	·6554217	36753	147	·3682701	14885	309
·41	·6590970	36602	150	·3667817	15190	305
·42	·6627573	36449	153	·3652627	15491	301
·43	·6664022	36293	156	·3637136	15787	296
·44	·6700314	36133	159	·3621349	16079	292
·45	·6736448	35971	162	·3605270	16367	288
·46	·6772419	35806	165	·3588903	16650	283
·47	·6808225	35638	168	·3572253	16928	278
·48	·6843863	35467	171	·3555325	17202	274
·49	·6879331	35294	173	·3538124	17470	269
·50	·6914625		176	·3520653		264

X	$P(X)$	δ +	δ^2 −
·50	·6914625	35118	176
·51	·6949743	34939	179
·52	·6984682	34758	181
·53	·7019440	34574	184
·54	·7054015	34388	186
·55	·7088403	34200	189
·56	·7122603	34009	191
·57	·7156612	33815	193
·58	·7190427	33620	196
·59	·7224047	33422	198
·60	·7257469	33222	200
·61	·7290691	33020	202
·62	·7323711	32816	204
·63	·7356527	32610	206
·64	·7389137	32402	208
·65	·7421539	32192	210
·66	·7453731	31980	212
·67	·7485711	31767	214
·68	·7517478	31551	215
·69	·7549029	31334	217
·70	·7580363	31116	219
·71	·7611479	30896	220
·72	·7642375	30674	222
·73	7673049	30451	223
·74	·7703500	30226	225
·75	·7733726	30001	226
·76	·7763727	29773	227
·77	·7793501	29545	228
·78	·7823046	29316	230
·79	·7852361	29085	231
·80	·7881446	28853	232
·81	·7910299	28620	233
·82	·7938919	28387	234
·83	·7967306	28152	235
·84	·7995458	27917	235
·85	·8023375	27680	236
·86	·8051055	27443	237
·87	·8078498	27205	238
·88	·8105703	26967	238
·89	·8132671	26728	239
·90	·8159399	26489	239
·91	·8185887	26249	240
·92	·8212136	26008	240
·93	·8238145	25768	241
·94	·8263912	25527	241
·95	·8289439	25285	241
·96	·8314724	25044	242
·97	·8339768	24802	242
·98	·8364569	24560	242
·99	·8389129	24318	242
1·00	·8413447		242

$$Z(X) = e^{-\frac{1}{2}X^2}/\sqrt{(2\pi)}, \quad P(X) = 1 - Q(X) = \int_{-\infty}^{X} Z(u)\, du.$$

The Normal Probability Function (continued)

$Z(X)$	δ −	δ^2 −
·3520653	17734	264
·3502919	17994	259
·3484925	18248	254
·3466677	18497	249
·3448180	18741	244
·3429439	18981	239
·3410458	19215	234
·3391243	19444	229
·3371799	19667	224
·3352132	19886	219
·3332246	20099	213
·3312147	20307	208
·3291840	20510	203
·3271330	20707	197
·3250623	20899	192
·3229724	21086	187
·3208638	21267	181
·3187371	21442	176
·3165929	21613	170
·3144317	21777	165
·3122539	21936	159
·3100603	22090	154
·3078513	22239	148
·3056274	22381	143
·3033893	22519	137
·3011374	22650	132
·2988724	22777	126
·2965948	22897	121
·2943050	23013	115
·2920038	23122	110
·2896916	23227	104
·2873689	23325	99
·2850364	23419	93
·2826945	23507	88
·2803438	23589	83
·2779849	23666	77
·2756182	23738	72
·2732444	23805	66
·2708640	23866	61
·2684774	23922	56
·2660852	23972	51
·2636880	24017	45
·2612863	24058	40
·2588805	24093	35
·2564713	24122	30
·2540591	24147	25
·2516443	24167	20
·2492277	24182	15
·2468095	24191	10
·2443904	24196	5
·2419707		0

X	$P(X)$	δ +	δ^2 −	$Z(X)$	δ −	δ^2 +
1·00	·8413447	24076	242	·2419707	24196	0
1·01	·8437524	23834	242	·2395511	24191	5
1·02	·8461358	23592	242	·2371320	24182	10
1·03	·8484950	23351	242	·2347138	24168	14
1·04	·8508300	23109	242	·2322970	24149	19
1·05	·8531409	22868	241	·2298821	24125	24
1·06	·8554277	22626	241	·2274696	24097	28
1·07	·8576903	22386	241	·2250599	24064	33
1·08	·8599289	22145	240	·2226535	24027	37
1·09	·8621434	21905	240	·2202508	23986	41
1·10	·8643339	21665	240	·2178522	23940	46
1·11	·8665005	21426	239	·2154582	23890	50
1·12	·8686431	21188	239	·2130691	23836	54
1·13	·8707619	20950	238	·2106856	23778	58
1·14	·8728568	20712	237	·2083078	23715	62
1·15	·8749281	20475	237	·2059363	23649	66
1·16	·8769756	20239	236	·2035714	23578	70
1·17	·8789995	20004	235	·2012135	23504	74
1·18	·8809999	19769	235	·1988631	23426	78
1·19	·8829768	19535	234	·1965205	23344	82
1·20	·8849303	19302	233	·1941861	23259	85
1·21	·8868606	19070	232	·1918602	23170	89
1·22	·8887676	18839	231	·1895432	23077	93
1·23	·8906514	18609	230	·1872354	22981	96
1·24	·8925123	18379	229	·1849373	22882	99
1·25	·8943502	18151	228	·1826491	22779	103
1·26	·8961653	17924	227	·1803712	22673	106
1·27	·8979577	17697	226	·1781038	22564	109
1·28	·8997274	17472	225	·1758474	22452	112
1·29	·9014747	17248	224	·1736022	22337	115
1·30	·9031995	17026	223	·1713686	22218	118
1·31	·9049021	16804	222	·1691468	22097	121
1·32	·9065825	16584	220	·1669370	21973	124
1·33	·9082409	16365	219	·1647397	21847	127
1·34	·9098773	16147	218	·1625551	21717	129
1·35	·9114920	15930	217	·1603833	21585	132
1·36	·9130850	15715	215	·1582248	21451	134
1·37	·9146565	15501	214	·1560797	21314	137
1·38	·9162067	15289	212	·1539483	21175	139
1·39	·9177356	15078	211	·1518308	21033	142
1·40	·9192433	14868	210	·1497275	20890	144
1·41	·9207302	14660	208	·1476385	20744	146
1·42	·9221962	14453	207	·1455641	20596	148
1·43	·9236415	14248	205	·1435046	20446	150
1·44	·9250663	14044	204	·1414600	20294	152
1·45	·9264707	13842	202	·1394306	20140	154
1·46	·9278550	13642	201	·1374165	19985	155
1·47	·9292191	13443	199	·1354181	19828	157
1·48	·9305634	13245	197	·1334353	19669	159
1·49	·9318879	13049	196	·1314684	19508	160
1·50	·9331928		194	·1295176		162

Note sign of second difference, δ^2.

The Normal Probability Function (continued)

X	$P(X)$	δ +	δ^2 −	$Z(X)$	δ −	δ^2 +
1·50	·9331928	12855	194	·1295176	19346	162
1·51	·9344783	12662	193	·1275830	19183	163
1·52	·9357445	12471	191	·1256646	19018	165
1·53	·9369916	12282	189	·1237628	18853	166
1·54	·9382198	12094	188	·1218775	18685	167
1·55	·9394292	11908	186	·1200090	18517	168
1·56	·9406201	11724	184	·1181573	18348	169
1·57	·9417924	11541	183	·1163225	18177	170
1·58	·9429466	11360	181	·1145048	18006	171
1·59	·9440826	11181	179	·1127042	17834	172
1·60	·9452007	11004	177	·1109208	17661	173
1·61	·9463011	10828	176	·1091548	17487	174
1·62	·9473839	10654	174	·1074061	17312	174
1·63	·9484493	10482	172	·1056748	17137	175
1·64	·9494974	10311	170	·1039611	16962	176
1·65	·9505285	10142	169	·1022649	16786	176
1·66	·9515428	9975	167	·1005864	16609	177
1·67	·9525403	9810	165	·0989255	16432	177
1·68	·9535213	9647	163	·0972823	16255	177
1·69	·9544860	9485	162	·0956568	16077	178
1·70	·9554345	9326	160	·0940491	15899	178
1·71	·9563671	9167	158	·0924591	15722	178
1·72	·9572838	9011	156	·0908870	15544	178
1·73	·9581849	8856	155	·0893326	15366	178
1·74	·9590705	8704	153	·0877961	15188	178
1·75	·9599408	8553	151	·0862773	15010	178
1·76	·9607961	8403	149	·0847764	14832	178
1·77	·9616364	8256	147	·0832932	14654	178
1·78	·9624620	8110	146	·0818278	14477	177
1·79	·9632730	7966	144	·0803801	14300	177
1·80	·9640697	7824	142	·0789502	14123	177
1·81	·9648521	7684	140	·0775379	13946	176
1·82	·9656205	7545	139	·0761433	13770	176
1·83	·9663750	7409	137	·0747663	13594	176
1·84	·9671159	7273	135	·0734068	13419	175
1·85	·9678432	7140	133	·0720649	13245	175
1·86	·9685572	7009	132	·0707404	13071	174
1·87	·9692581	6879	130	·0694333	12897	173
1·88	·9699460	6751	128	·0681436	12725	173
1·89	·9706210	6624	126	·0668711	12553	172
1·90	·9712834	6500	125	·0656158	12382	171
1·91	·9719334	6377	123	·0643777	12211	170
1·92	·9725711	6255	121	·0631566	12041	170
1·93	·9731966	6136	120	·0619524	11873	169
1·94	·9738102	6018	118	·0607652	11705	168
1·95	·9744119	5902	116	·0595947	11538	167
1·96	·9750021	5787	115	·0584409	11372	166
1·97	·9755808	5674	113	·0573038	11206	165
1·98	·9761482	5563	111	·0561831	11042	164
1·99	·9767045	5453	110	·0550789	10879	163
2·00	·9772499		108	·0539910		162

X	$P(X)$	δ +	δ^2 −
2·00	·9772499	5345	108
2·01	·9777844	5239	106
2·02	·9783083	5134	105
2·03	·9788217	5031	103
2·04	·9793248	4929	102
2·05	·9798178	4829	100
2·06	·9803007	4731	98
2·07	·9807738	4634	97
2·08	·9812372	4539	95
2·09	·9816911	4445	94
2·10	·9821356	4352	92
2·11	·9825708	4262	91
2·12	·9829970	4172	89
2·13	·9834142	4084	88
2·14	·9838226	3998	86
2·15	·9842224	3913	85
2·16	·9846137	3829	84
2·17	·9849966	3747	82
2·18	·9853713	3666	81
2·19	·9857379	3587	79
2·20	·9860966	3509	78
2·21	·9864474	3432	77
2·22	·9867906	3357	75
2·23	·9871263	3283	74
2·24	·9874545	3210	73
2·25	·9877755	3138	71
2·26	·9880894	3068	70
2·27	·9883962	2999	69
2·28	·9886962	2932	68
2·29	·9889893	2865	66
2·30	·9892759	2800	65
2·31	·9895559	2736	64
2·32	·9898296	2674	63
2·33	·9900969	2612	62
2·34	·9903581	2552	60
2·35	·9906133	2492	59
2·36	·9908625	2434	58
2·37	·9911060	2377	57
2·38	·9913437	2321	56
2·39	·9915758	2267	55
2·40	·9918025	2213	54
2·41	·9920237	2160	53
2·42	·9922397	2108	52
2·43	·9924506	2058	51
2·44	·9926564	2008	50
2·45	·9928572	1960	49
2·46	·9930531	1912	48
2·47	·9932443	1865	47
2·48	·9934309	1820	46
2·49	·9936128	1775	45
2·50	·9937903		44

$$Z(X)=e^{-\frac{1}{2}X^2}/\sqrt{(2\pi)},\quad P(X)=1-Q(X)=\int_{-\infty}^{X} Z(u)\,du.$$

The Normal Probability Function (continued)

$Z(X)$	δ −	δ^2 +
·0539910	10717	162
·0529192	10557	161
·0518636	10397	160
·0508239	10238	159
·0498001	10081	157
·0487920	9924	156
·0477996	9769	155
·0468226	9616	154
·0458611	9463	153
·0449148	9312	151
·0439836	9162	150
·0430674	9013	149
·0421661	8866	147
·0412795	8720	146
·0404076	8575	145
·0395500	8432	143
·0387069	8290	142
·0378779	8149	140
·0370629	8010	139
·0362619	7873	138
·0354746	7737	136
·0347009	7602	135
·0339408	7468	133
·0331939	7337	132
·0324603	7206	130
·0317397	7077	129
·0310319	6950	127
·0303370	6824	126
·0296546	6699	125
·0289847	6576	123
·0283270	6455	122
·0276816	6335	120
·0270481	6216	119
·0264265	6099	117
·0258166	5984	116
·0252182	5870	114
·0246313	5757	113
·0240556	5646	111
·0234910	5536	110
·0229374	5428	108
·0223945	5322	107
·0218624	5217	105
·0213407	5113	104
·0208294	5011	102
·0203284	4910	101
·0198374	4811	99
·0193563	4713	98
·0188850	4617	96
·0184233	4522	95
·0179711	4428	93
·0175283		92

X	$P(X)$	δ +	δ^2 −	$Z(X)$	δ −	δ^2 +
2·50	·9937903	1731	44	·0175283	4336	92
2·51	·9939634	1688	43	·0170947	4246	91
2·52	·9941323	1646	42	·0166701	4157	89
2·53	·9942969	1605	41	·0162545	4069	88
2·54	·9944574	1565	40	·0158476	3982	86
2·55	·9946139	1525	39	·0154493	3897	85
2·56	·9947664	1487	39	·0150596	3814	84
2·57	·9949151	1449	38	·0146782	3731	82
2·58	·9950600	1412	37	·0143051	3650	81
2·59	·9952012	1376	36	·0139401	3571	80
2·60	·9953388	1341	35	·0135830	3493	78
2·61	·9954729	1306	35	·0132337	3416	77
2·62	·9956035	1272	34	·0128921	3340	76
2·63	·9957308	1239	33	·0125581	3266	74
2·64	·9958547	1207	32	·0122315	3193	73
2·65	·9959754	1176	32	·0119122	3121	72
2·66	·9960930	1145	31	·0116001	3051	70
2·67	·9962074	1115	30	·0112951	2981	69
2·68	·9963189	1085	29	·0109969	2913	68
2·69	·9964274	1056	29	·0107056	2847	67
2·70	·9965330	1028	28	·0104209	2781	66
2·71	·9966358	1001	27	·0101428	2717	64
2·72	·9967359	974	27	·0098712	2654	63
2·73	·9968333	948	26	·0096058	2592	62
2·74	·9969280	922	26	·0093466	2531	61
2·75	·9970202	897	25	·0090936	2471	60
2·76	·9971099	873	24	·0088465	2413	59
2·77	·9971972	849	24	·0086052	2355	57
2·78	·9972821	825	23	·0083697	2299	56
2·79	·9973646	803	23	·0081398	2244	55
2·80	·9974449	781	22	·0079155	2189	54
2·81	·9975229	759	22	·0076965	2136	53
2·82	·9975988	738	21	·0074829	2084	52
2·83	·9976726	717	21	·0072744	2033	51
2·84	·9977443	697	20	·0070711	1983	50
2·85	·9978140	678	20	·0068728	1934	49
2·86	·9978818	658	19	·0066793	1886	48
2·87	·9979476	640	19	·0064907	1839	47
2·88	·9980116	622	18	·0063067	1793	46
2·89	·9980738	604	18	·0061274	1748	45
2·90	·9981342	587	17	·0059525	1704	44
2·91	·9981929	570	17	·0057821	1661	43
2·92	·9982498	553	16	·0056160	1619	42
2·93	·9983052	537	16	·0054541	1578	41
2·94	·9983589	522	16	·0052963	1537	40
2·95	·9984111	507	15	·0051426	1497	40
2·96	·9984618	492	15	·0049929	1459	39
2·97	·9985110	478	14	·0048470	1421	38
2·98	·9985588	464	14	·0047050	1384	37
2·99	·9986051	450	14	·0045666	1347	36
3·00	·9986501		13	·0044318		35

Note sign of second difference, δ^2.

The Normal Probability Function (continued)

X	$P(X)$	δ +	δ^2 −	$Z(X)$	δ −	δ^2 +
3·00	·9986501	437	13	·0044318	1312	35
3·01	·9986938	424	13	·0043007	1277	35
3·02	·9987361	411	13	·0041729	1243	34
3·03	·9987772	399	12	·0040486	1210	33
3·04	·9988171	387	12	·0039276	1178	32
3·05	·9988558	375	12	·0038098	1146	32
3·06	·9988933	364	11	·0036951	1115	31
3·07	·9989297	353	11	·0035836	1085	30
3·08	·9989650	342	11	·0034751	1056	29
3·09	·9989992	332	10	·0033695	1027	29
3·10	·9990324	322	10	·0032668	999	28
3·11	·9990646	312	10	·0031669	971	27
3·12	·9990957	302	10	·0030698	944	27
3·13	·9991260	293	9	·0029754	918	26
3·14	·9991553	284	9	·0028835	893	26
3·15	·9991836	276	9	·0027943	868	25
3·16	·9992112	267	9	·0027075	843	24
3·17	·9992378	258	8	·0026231	820	24
3·18	·9992636	250	8	·0025412	797	23
3·19	·9992886	242	8	·0024615	774	23
3·20	·9993129	235	8	·0023841	752	22
3·21	·9993363	227	7	·0023089	731	21
3·22	·9993590	220	7	·0022358	710	21
3·23	·9993810	213	7	·0021649	689	20
3·24	·9994024	206	7	·0020960	669	20
3·25	·9994230	200	7	·0020290	650	19
3·26	·9994429	193	6	·0019641	631	19
3·27	·9994623	187	6	·0019010	612	18
3·28	·9994810	181	6	·0018397	595	18
3·29	·9994991	175	6	·0017803	577	17
3·30	·9995166	169	6	·0017226	560	17
3·31	·9995335	164	6	·0016666	543	17
3·32	·9995499	159	5	·0016122	527	16
3·33	·9995658	153	5	·0015595	512	16
3·34	·9995811	148	5	·0015084	496	15
3·35	·9995959	143	5	·0014587	481	15
3·36	·9996103	139	5	·0014106	467	15
3·37	·9996242	134	5	·0013639	453	14
3·38	·9996376	130	4	·0013187	439	14
3·39	·9996505	125	4	·0012748	426	13
3·40	·9996631	121	4	·0012322	413	13
3·41	·9996752	117	4	·0011910	400	13
3·42	·9996869	113	4	·0011510	388	12
3·43	·9996982	109	4	·0011122	376	12
3·44	·9997091	106	4	·0010747	364	12
3·45	·9997197	102	4	·0010383	353	11
3·46	·9997299	99	3	·0010030	342	11
3·47	·9997398	95	3	·0009689	331	11
3·48	·9997493	92	3	·0009358	320	10
3·49	·9997585	89	3	·0009037	310	10
3·50	·9997674		3	·0008727		10

X	$P(X)$	δ +	δ^2 −
3·50	·9997674	86	3
3·51	·9997759	83	3
3·52	·9997842	80	3
3·53	·9997922	77	3
3·54	·9997999	74	3
3·55	·9998074	72	3
3·56	·9998146	69	3
3·57	·9998215	67	2
3·58	·9998282	65	2
3·59	·9998347	62	2
3·60	9998409	60	2
3·61	9998469	58	2
3·62	·9998527	56	2
3·63	·9998583	54	2
3·64	·9998637	52	2
3·65	·9998689	50	2
3·66	·9998739	48	2
3·67	·9998787	47	2
3·68	·9998834	45	2
3·69	·9998879	43	2
3·70	·9998922	42	2
3·71	·9998964	40	2
3·72	·9999004	39	1
3·73	·9999043	37	1
3·74	·9999080	36	1
3·75	·9999116	35	1
3·76	·9999150	33	1
3·77	·9999184	32	1
3·78	·9999216	31	1
3·79	·9999247	30	1
3·80	·9999277	29	1
3·81	·9999305	28	1
3·82	·9999333	27	1
3·83	·9999359	26	1
3·84	·9999385	25	1
3·85	·9999409	24	1
3·86	·9999433	23	1
3·87	·9999456	22	1
3·88	·9999478	21	1
3·89	·9999499	20	1
3·90	·9999519	19	1
3·91	·9999539	19	1
3·92	·9999557	18	1
3·93	·9999575	17	1
3·94	·9999593	17	1
3·95	·9999609	16	1
3·96	·9999625	16	1
3·97	·9999641	15	1
3·98	·9999655	14	1
3·99	·9999670	14	1
4·00	·9999683		1

$$Z(X)=e^{-\frac{1}{2}X^2}/\sqrt{(2\pi)},\quad P(X)=1-Q(X)=\int_{-\infty}^{X} Z(u)\,du.$$

The Normal Probability Function (continued)

$Z(X)$	δ −	δ^2 +
·0008727	301	10
·0008426	291	10
·0008135	282	9
·0007853	273	9
·0007581	264	9
·0007317	256	8
·0007061	247	8
·0006814	239	8
·0006575	232	8
·0006343	224	8
·0006119	217	7
·0005902	210	7
·0005693	203	7
·0005490	196	7
·0005294	189	6
·0005105	183	6
·0004921	177	6
·0004744	171	6
·0004573	165	6
·0004408	160	6
·0004248	155	5
·0004093	149	5
·0003944	144	5
·0003800	139	5
·0003661	135	5
·0003526	130	5
·0003396	125	4
·0003271	121	4
·0003149	117	4
·0003032	113	4
·0002919	109	4
·0002810	105	4
·0002705	102	4
·0002604	98	4
·0002506	95	3
·0002411	91	3
·0002320	88	3
·0002232	85	3
·0002147	82	3
·0002065	79	3
·0001987	76	3
·0001910	73	3
·0001837	71	3
·0001766	68	3
·0001698	66	2
·0001633	63	2
·0001569	61	2
·0001508	59	2
·0001449	57	2
·0001393	55	2
·0001338		2

X	$P(X)$	δ +	δ^2 −	$Z(X)$	δ −	δ^2 +
4·00	·9999683	13	1	·0001338	53	2
4·01	·9999696	13	1	·0001286	51	2
4·02	·9999709	12	0	·0001235	49	2
4·03	·9999721	12		·0001186	47	2
4·04	·9999733	11		·0001140	45	2
4·05	·9999744	11		·0001094	43	2
4·06	·9999755	10		·0001051	42	2
4·07	·9999765	10		·0001009	40	2
4·08	·9999775	9		·0000969	39	2
4·09	·9999784	9		·0000930	37	1
4·10	·9999793	9		·0000893	36	1
4·11	·9999802	8		·0000857	35	1
4·12	·9999811	8		·0000822	33	1
4·13	·9999819	8		·0000789	32	1
4·14	·9999826	7		·0000757	31	1
4·15	·9999834	7		·0000726	30	1
4·16	·9999841	7		·0000697	28	1
4·17	·9999848	7		·0000668	27	1
4·18	·9999854	6		·0000641	26	1
4·19	·9999861	6		·0000615	25	1
4·20	·9999867	6		·0000589	24	1
4·21	·9999872	6		·0000565	23	1
4·22	·9999878	5		·0000542	22	1
4·23	·9999883	5		·0000519	22	1
4·24	·9999888	5		·0000498	21	1
4·25	·9999893	5		·0000477	20	1
4·26	·9999898	4		·0000457	19	1
4·27	·9999902	4		·0000438	18	1
4·28	·9999907	4		·0000420	18	1
4·29	·9999911	4		·0000402	17	1
4·30	·9999915	4		·0000385	16	1
4·31	·9999918	4		·0000369	16	1
4·32	·9999922	3		·0000354	15	1
4·33	·9999925	3		·0000339	14	1
4·34	·9999929	3		·0000324	14	1
4·35	·9999932	3		·0000310	13	1
4·36	·9999935	3		·0000297	13	1
4·37	·9999938	3		·0000284	12	1
4·38	·9999941	3		·0000272	12	0
4·39	·9999943	3		·0000261	11	
4·40	·9999946	2		·0000249	11	
4·41	·9999948	2		·0000239	10	
4·42	·9999951	2		·0000228	10	
4·43	·9999953	2		·0000218	9	
4·44	·9999955	2		·0000209	9	
4·45	·9999957	2		·0000200	9	
4·46	·9999959	2		·0000191	8	
4·47	·9999961	2		·0000183	8	
4·48	·9999963	2		·0000175	8	
4·49	·9999964	2		·0000167	7	
4·50	·9999966			·0000160		

Note sign of second difference, δ^2.

The Normal Probability Function (continued)

X	$P(X)$*	$Z(X)$*	X	$P(X)$*	$Z(X)$*	X	$P(X)$*	$Z(X)$*
4·50	66023	159837	5·00	97133	14867	5·50	99810	1077
4·51	67586	152797	5·01	97278	14141	5·51	99821	1019
4·52	69080	146051	5·02	97416	13450	5·52	99831	965
4·53	70508	139590	5·03	97548	12791	5·53	99840	913
4·54	71873	133401	5·04	97672	12162	5·54	99849	864
4·55	73177	127473	5·05	97791	11564	5·55	99857	817
4·56	74423	121797	5·06	97904	10994	5·56	99865	773
4·57	75614	116362	5·07	98011	10451	5·57	99873	731
4·58	76751	111159	5·08	98113	9934	5·58	99880	691
4·59	77838	106177	5·09	98210	9441	5·59	99886	654
4·60	78875	101409	5·10	98302	8972	5·60	99893	618
4·61	79867	96845	5·11	98389	8526	5·61	99899	585
4·62	80813	92477	5·12	98472	8101	5·62	99905	553
4·63	81717	88297	5·13	98551	7696	5·63	99910	522
4·64	82580	84298	5·14	98626	7311	5·64	99915	494
4·65	83403	80472	5·15	98698	6944	5·65	99920	467
4·66	84190	76812	5·16	98765	6595	5·66	99924	441
4·67	84940	73311	5·17	98830	6263	5·67	99929	417
4·68	85656	69962	5·18	98891	5947	5·68	99933	394
4·69	86340	66760	5·19	98949	5647	5·69	99936	372
4·70	86992	63698	5·20	99004	5361	5·70	99940	351
4·71	87614	60771	5·21	99056	5089	5·71	99944	332
4·72	88208	57972	5·22	99105	4831	5·72	99947	313
4·73	88774	55296	5·23	99152	4585	5·73	99950	296
4·74	89314	52739	5·24	99197	4351	5·74	99953	280
4·75	89829	50295	5·25	99240	4128	5·75	99955	264
4·76	90320	47960	5·26	99280	3917	5·76	99958	249
4·77	90789	45728	5·27	99318	3716	5·77	99960	235
4·78	91235	43596	5·28	99354	3525	5·78	99963	222
4·79	91661	41559	5·29	99388	3344	5·79	99965	210
4·80	92067	39613	5·30	99421	3171	5·80	99967	198
4·81	92453	37755	5·31	99452	3007	5·81	99969	187
4·82	92822	35980	5·32	99481	2852	5·82	99971	176
4·83	93173	34285	5·33	99509	2704	5·83	99972	166
4·84	93508	32667	5·34	99535	2563	5·84	99974	157
4·85	93827	31122	5·35	99560	2430	5·85	99975	148
4·86	94131	29647	5·36	99584	2303	5·86	99977	139
4·87	94420	28239	5·37	99606	2183	5·87	99978	131
4·88	94696	26895	5·38	99628	2069	5·88	99979	124
4·89	94958	25613	5·39	99648	1960	5·89	99981	117
4·90	95208	24390	5·40	99667	1857	5·90	99982	110
4·91	95446	23222	5·41	99685	1760	5·91	99983	104
4·92	95673	22108	5·42	99702	1667	5·92	99984	98
4·93	95889	21046	5·43	99718	1579	5·93	99985	92
4·94	96094	20033	5·44	99734	1495	5·94	99986	87
4·95	96289	19068	5·45	99748	1416	5·95	99987	82
4·96	96475	18144	5·46	99762	1341	5·96	99987	77
4·97	96652	17265	5·47	99775	1270	5·97	99988	73
4·98	96821	16428	5·48	99787	1202	5·98	99989	68
4·99	96981	15629	5·49	99799	1138	5·99	99990	65
						6·00	99990	61

$$Z(X)=e^{-\frac{1}{2}X^2}/\sqrt{(2\pi)}, \quad P(X)=1-Q(X)=\int_{-\infty}^{X} Z(u)\,du.$$

* The entries for $P(X)$ and $Z(X)$ on this page are given to 10 decimal places; thus 0·99999 should be prefixed to each entry for $P(X)$ and a decimal point, followed by four, five, ..., eight zeros, as appropriate, to $Z(X)$.

This table was reprinted from Biometrika Tables for Statisticians, Vol. 1, 3rd Edition, Table 1, with the permission of the Biometrika Trustees.

Percentage Points of the F-distribution (Variance Ratio)

Upper 25% points

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5·83	7·50	8·20	8·58	8·82	8·98	9·10	9·19	9·26	9·32	9·41	9·49	9·58	9·63	9·67	9·71	9·76	9·80	9·85
2	2·57	3·00	3·15	3·23	3·28	3·31	3·34	3·35	3·37	3·38	3·39	3·41	3·43	3·43	3·44	3·45	3·46	3·47	3·48
3	2·02	2·28	2·36	2·39	2·41	2·42	2·43	2·44	2·44	2·44	2·45	2·46	2·46	2·46	2·47	2·47	2·47	2·47	2·47
4	1·81	2·00	2·05	2·06	2·07	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08	2·08
5	1·69	1·85	1·88	1·89	1·89	1·89	1·89	1·89	1·89	1·89	1·89	1·89	1·88	1·88	1·88	1·88	1·87	1·87	1·87
6	1·62	1·76	1·78	1·79	1·79	1·78	1·78	1·78	1·77	1·77	1·77	1·76	1·76	1·75	1·75	1·75	1·74	1·74	1·74
7	1·57	1·70	1·72	1·72	1·71	1·71	1·70	1·70	1·69	1·69	1·68	1·68	1·67	1·67	1·66	1·66	1·65	1·65	1·65
8	1·54	1·66	1·67	1·66	1·66	1·65	1·64	1·64	1·63	1·63	1·62	1·62	1·61	1·60	1·60	1·59	1·59	1·58	1·58
9	1·51	1·62	1·63	1·63	1·62	1·61	1·60	1·60	1·59	1·59	1·58	1·57	1·56	1·56	1·55	1·54	1·54	1·53	1·53
10	1·49	1·60	1·60	1·59	1·59	1·58	1·57	1·56	1·56	1·55	1·54	1·53	1·52	1·52	1·51	1·51	1·50	1·49	1·48
11	1·47	1·58	1·58	1·57	1·56	1·55	1·54	1·53	1·53	1·52	1·51	1·50	1·49	1·49	1·48	1·47	1·47	1·46	1·45
12	1·46	1·56	1·56	1·55	1·54	1·53	1·52	1·51	1·51	1·50	1·49	1·48	1·47	1·46	1·45	1·45	1·44	1·43	1·42
13	1·45	1·55	1·55	1·53	1·52	1·51	1·50	1·49	1·49	1·48	1·47	1·46	1·45	1·44	1·43	1·42	1·42	1·41	1·40
14	1·44	1·53	1·53	1·52	1·51	1·50	1·49	1·48	1·47	1·46	1·45	1·44	1·43	1·42	1·41	1·41	1·40	1·39	1·38
15	1·43	1·52	1·52	1·51	1·49	1·48	1·47	1·46	1·46	1·45	1·44	1·43	1·41	1·41	1·40	1·39	1·38	1·37	1·36
16	1·42	1·51	1·51	1·50	1·48	1·47	1·46	1·45	1·44	1·44	1·43	1·41	1·40	1·39	1·38	1·37	1·36	1·35	1·34
17	1·42	1·51	1·50	1·49	1·47	1·46	1·45	1·44	1·43	1·43	1·41	1·40	1·39	1·38	1·37	1·36	1·35	1·34	1·33
18	1·41	1·50	1·49	1·48	1·46	1·45	1·44	1·43	1·42	1·42	1·40	1·39	1·38	1·37	1·36	1·35	1·34	1·33	1·32
19	1·41	1·49	1·49	1·47	1·46	1·44	1·43	1·42	1·41	1·41	1·40	1·38	1·37	1·36	1·35	1·34	1·33	1·32	1·30
20	1·40	1·49	1·48	1·47	1·45	1·44	1·43	1·42	1·41	1·40	1·39	1·37	1·36	1·35	1·34	1·33	1·32	1·31	1·29
21	1·40	1·48	1·48	1·46	1·44	1·43	1·42	1·41	1·40	1·39	1·38	1·37	1·35	1·34	1·33	1·32	1·31	1·30	1·28
22	1·40	1·48	1·47	1·45	1·44	1·42	1·41	1·40	1·39	1·39	1·37	1·36	1·34	1·33	1·32	1·31	1·30	1·29	1·28
23	1·39	1·47	1·47	1·45	1·43	1·42	1·41	1·40	1·39	1·38	1·37	1·35	1·34	1·33	1·32	1·31	1·30	1·28	1·27
24	1·39	1·47	1·46	1·44	1·43	1·41	1·40	1·39	1·38	1·38	1·36	1·35	1·33	1·32	1·31	1·30	1·29	1·28	1·26
25	1·39	1·47	1·46	1·44	1·42	1·41	1·40	1·39	1·38	1·37	1·36	1·34	1·33	1·32	1·31	1·29	1·28	1·27	1·25
26	1·38	1·46	1·45	1·44	1·42	1·41	1·39	1·38	1·37	1·37	1·35	1·34	1·32	1·31	1·30	1·29	1·28	1·26	1·25
27	1·38	1·46	1·45	1·43	1·42	1·40	1·39	1·38	1·37	1·36	1·35	1·33	1·32	1·31	1·30	1·28	1·27	1·26	1·24
28	1·38	1·46	1·45	1·43	1·41	1·40	1·39	1·38	1·37	1·36	1·34	1·33	1·31	1·30	1·29	1·28	1·27	1·25	1·24
29	1·38	1·45	1·45	1·43	1·41	1·40	1·38	1·37	1·36	1·35	1·34	1·32	1·31	1·30	1·29	1·27	1·26	1·25	1·23
30	1·38	1·45	1·44	1·42	1·41	1·39	1·38	1·37	1·36	1·35	1·34	1·32	1·30	1·29	1·28	1·27	1·26	1·24	1·23
40	1·36	1·44	1·42	1·40	1·39	1·37	1·36	1·35	1·34	1·33	1·31	1·30	1·28	1·26	1·25	1·24	1·22	1·21	1·19
60	1·35	1·42	1·41	1·38	1·37	1·35	1·33	1·32	1·31	1·30	1·29	1·27	1·25	1·24	1·22	1·21	1·19	1·17	1·15
120	1·34	1·40	1·39	1·37	1·35	1·33	1·31	1·30	1·29	1·28	1·26	1·24	1·22	1·21	1·19	1·18	1·16	1·13	1·10
∞	1·32	1·39	1·37	1·35	1·33	1·31	1·29	1·28	1·27	1·25	1·24	1·22	1·19	1·18	1·16	1·14	1·12	1·08	1·00

$F = \dfrac{s_1^2}{s_2^2} = \dfrac{S_1}{\nu_1}\Big/\dfrac{S_2}{\nu_2}$, where $s_1^2 = S_1/\nu_1$ and $s_2^2 = S_2/\nu_2$ are independent mean squares estimating a common variance σ^2 and based on ν_1 and ν_2 degrees of freedom, respectively.

Percentage Points of the F-distribution (Variance Ratio) (continued)

Upper 10% points

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39·86	49·50	53·59	55·83	57·24	58·20	58·91	59·44	59·86	60·19	60·71	61·22	61·74	62·00	62·26	62·53	62·79	63·06	63·33
2	8·53	9·00	9·16	9·24	9·29	9·33	9·35	9·37	9·38	9·39	9·41	9·42	9·44	9·45	9·46	9·47	9·47	9·48	9·49
3	5·54	5·46	5·39	5·34	5·31	5·28	5·27	5·25	5·24	5·23	5·22	5·20	5·18	5·18	5·17	5·16	5·15	5·14	5·13
4	4·54	4·32	4·19	4·11	4·05	4·01	3·98	3·95	3·94	3·92	3·90	3·87	3·84	3·83	3·82	3·80	3·79	3·78	3·76
5	4·06	3·78	3·62	3·52	3·45	3·40	3·37	3·34	3·32	3·30	3·27	3·24	3·21	3·19	3·17	3·16	3·14	3·12	3·10
6	3·78	3·46	3·29	3·18	3·11	3·05	3·01	2·98	2·96	2·94	2·90	2·87	2·84	2·82	2·80	2·78	2·76	2·74	2·72
7	3·59	3·26	3·07	2·96	2·88	2·83	2·78	2·75	2·72	2·70	2·67	2·63	2·59	2·58	2·56	2·54	2·51	2·49	2·47
8	3·46	3·11	2·92	2·81	2·73	2·67	2·62	2·59	2·56	2·54	2·50	2·46	2·42	2·40	2·38	2·36	2·34	2·32	2·29
9	3·36	3·01	2·81	2·69	2·61	2·55	2·51	2·47	2·44	2·42	2·38	2·34	2·30	2·28	2·25	2·23	2·21	2·18	2·16
10	3·29	2·92	2·73	2·61	2·52	2·46	2·41	2·38	2·35	2·32	2·28	2·24	2·20	2·18	2·16	2·13	2·11	2·08	2·06
11	3·23	2·86	2·66	2·54	2·45	2·39	2·34	2·30	2·27	2·25	2·21	2·17	2·12	2·10	2·08	2·05	2·03	2·00	1·97
12	3·18	2·81	2·61	2·48	2·39	2·33	2·28	2·24	2·21	2·19	2·15	2·10	2·06	2·04	2·01	1·99	1·96	1·93	1·90
13	3·14	2·76	2·56	2·43	2·35	2·28	2·23	2·20	2·16	2·14	2·10	2·05	2·01	1·98	1·96	1·93	1·90	1·88	1·85
14	3·10	2·73	2·52	2·39	2·31	2·24	2·19	2·15	2·12	2·10	2·05	2·01	1·96	1·94	1·91	1·89	1·86	1·83	1·80
15	3·07	2·70	2·49	2·36	2·27	2·21	2·16	2·12	2·09	2·06	2·02	1·97	1·92	1·90	1·87	1·85	1·82	1·79	1·76
16	3·05	2·67	2·46	2·33	2·24	2·18	2·13	2·09	2·06	2·03	1·99	1·94	1·89	1·87	1·84	1·81	1·78	1·75	1·72
17	3·03	2·64	2·44	2·31	2·22	2·15	2·10	2·06	2·03	2·00	1·96	1·91	1·86	1·84	1·81	1·78	1·75	1·72	1·69
18	3·01	2·62	2·42	2·29	2·20	2·13	2·08	2·04	2·00	1·98	1·93	1·89	1·84	1·81	1·78	1·75	1·72	1·69	1·66
19	2·99	2·61	2·40	2·27	2·18	2·11	2·06	2·02	1·98	1·96	1·91	1·86	1·81	1·79	1·76	1·73	1·70	1·67	1·63
20	2·97	2·59	2·38	2·25	2·16	2·09	2·04	2·00	1·96	1·94	1·89	1·84	1·79	1·77	1·74	1·71	1·68	1·64	1·61
21	2·96	2·57	2·36	2·23	2·14	2·08	2·02	1·98	1·95	1·92	1·87	1·83	1·78	1·75	1·72	1·69	1·66	1·62	1·59
22	2·95	2·56	2·35	2·22	2·13	2·06	2·01	1·97	1·93	1·90	1·86	1·81	1·76	1·73	1·70	1·67	1·64	1·60	1·57
23	2·94	2·55	2·34	2·21	2·11	2·05	1·99	1·95	1·92	1·89	1·84	1·80	1·74	1·72	1·69	1·66	1·62	1·59	1·55
24	2·93	2·54	2·33	2·19	2·10	2·04	1·98	1·94	1·91	1·88	1·83	1·78	1·73	1·70	1·67	1·64	1·61	1·57	1·53
25	2·92	2·53	2·32	2·18	2·09	2·02	1·97	1·93	1·89	1·87	1·82	1·77	1·72	1·69	1·66	1·63	1·59	1·56	1·52
26	2·91	2·52	2·31	2·17	2·08	2·01	1·96	1·92	1·88	1·86	1·81	1·76	1·71	1·68	1·65	1·61	1·58	1·54	1·50
27	2·90	2·51	2·30	2·17	2·07	2·00	1·95	1·91	1·87	1·85	1·80	1·75	1·70	1·67	1·64	1·60	1·57	1·53	1·49
28	2·89	2·50	2·29	2·16	2·06	2·00	1·94	1·90	1·87	1·84	1·79	1·74	1·69	1·66	1·63	1·59	1·56	1·52	1·48
29	2·89	2·50	2·28	2·15	2·06	1·99	1·93	1·89	1·86	1·83	1·78	1·73	1·68	1·65	1·62	1·58	1·55	1·51	1·47
30	2·88	2·49	2·28	2·14	2·05	1·98	1·93	1·88	1·85	1·82	1·77	1·72	1·67	1·64	1·61	1·57	1·54	1·50	1·46
40	2·84	2·44	2·23	2·09	2·00	1·93	1·87	1·83	1·79	1·76	1·71	1·66	1·61	1·57	1·54	1·51	1·47	1·42	1·38
60	2·79	2·39	2·18	2·04	1·95	1·87	1·82	1·77	1·74	1·71	1·66	1·60	1·54	1·51	1·48	1·44	1·40	1·35	1·29
120	2·75	2·35	2·13	1·99	1·90	1·82	1·77	1·72	1·68	1·65	1·60	1·55	1·48	1·45	1·41	1·37	1·32	1·26	1·19
∞	2·71	2·30	2·08	1·94	1·85	1·77	1·72	1·67	1·63	1·60	1·55	1·49	1·42	1·38	1·34	1·30	1·24	1·17	1·00

$F = \dfrac{s_1^2}{s_2^2} = \dfrac{S_1}{\nu_1}\Big/\dfrac{S_2}{\nu_2}$, where $s_1^2 = S_1/\nu_1$ and $s_2^2 = S_2/\nu_2$ are independent mean squares estimating a common variance σ^2 and based on ν_1 and ν_2 degrees of freedom, respectively.

Percentage Points of the F-distribution (Variance Ratio) (continued)

Upper 5% points

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161·4	199·5	215·7	224·6	230·2	234·0	236·8	238·9	240·5	241·9	243·9	245·9	248·0	249·1	250·1	251·1	252·2	253·3	254·3
2	18·51	19·00	19·16	19·25	19·30	19·33	19·35	19·37	19·38	19·40	19·41	19·43	19·45	19·45	19·46	19·47	19·48	19·49	19·50
3	10·13	9·55	9·28	9·12	9·01	8·94	8·89	8·85	8·81	8·79	8·74	8·70	8·66	8·64	8·62	8·59	8·57	8·55	8·53
4	7·71	6·94	6·59	6·39	6·26	6·16	6·09	6·04	6·00	5·96	5·91	5·86	5·80	5·77	5·75	5·72	5·69	5·66	5·63
5	6·61	5·79	5·41	5·19	5·05	4·95	4·88	4·82	4·77	4·74	4·68	4·62	4·56	4·53	4·50	4·46	4·43	4·40	4·36
6	5·99	5·14	4·76	4·53	4·39	4·28	4·21	4·15	4·10	4·06	4·00	3·94	3·87	3·84	3·81	3·77	3·74	3·70	3·67
7	5·59	4·74	4·35	4·12	3·97	3·87	3·79	3·73	3·68	3·64	3·57	3·51	3·44	3·41	3·38	3·34	3·30	3·27	3·23
8	5·32	4·46	4·07	3·84	3·69	3·58	3·50	3·44	3·39	3·35	3·28	3·22	3·15	3·12	3·08	3·04	3·01	2·97	2·93
9	5·12	4·26	3·86	3·63	3·48	3·37	3·29	3·23	3·18	3·14	3·07	3·01	2·94	2·90	2·86	2·83	2·79	2·75	2·71
10	4·96	4·10	3·71	3·48	3·33	3·22	3·14	3·07	3·02	2·98	2·91	2·85	2·77	2·74	2·70	2·66	2·62	2·58	2·54
11	4·84	3·98	3·59	3·36	3·20	3·09	3·01	2·95	2·90	2·85	2·79	2·72	2·65	2·61	2·57	2·53	2·49	2·45	2·40
12	4·75	3·89	3·49	3·26	3·11	3·00	2·91	2·85	2·80	2·75	2·69	2·62	2·54	2·51	2·47	2·43	2·38	2·34	2·30
13	4·67	3·81	3·41	3·18	3·03	2·92	2·83	2·77	2·71	2·67	2·60	2·53	2·46	2·42	2·38	2·34	2·30	2·25	2·21
14	4·60	3·74	3·34	3·11	2·96	2·85	2·76	2·70	2·65	2·60	2·53	2·46	2·39	2·35	2·31	2·27	2·22	2·18	2·13
15	4·54	3·68	3·29	3·06	2·90	2·79	2·71	2·64	2·59	2·54	2·48	2·40	2·33	2·29	2·25	2·20	2·16	2·11	2·07
16	4·49	3·63	3·24	3·01	2·85	2·74	2·66	2·59	2·54	2·49	2·42	2·35	2·28	2·24	2·19	2·15	2·11	2·06	2·01
17	4·45	3·59	3·20	2·96	2·81	2·70	2·61	2·55	2·49	2·45	2·38	2·31	2·23	2·19	2·15	2·10	2·06	2·01	1·96
18	4·41	3·55	3·16	2·93	2·77	2·66	2·58	2·51	2·46	2·41	2·34	2·27	2·19	2·15	2·11	2·06	2·02	1·97	1·92
19	4·38	3·52	3·13	2·90	2·74	2·63	2·54	2·48	2·42	2·38	2·31	2·23	2·16	2·11	2·07	2·03	1·98	1·93	1·88
20	4·35	3·49	3·10	2·87	2·71	2·60	2·51	2·45	2·39	2·35	2·28	2·20	2·12	2·08	2·04	1·99	1·95	1·90	1·84
21	4·32	3·47	3·07	2·84	2·68	2·57	2·49	2·42	2·37	2·32	2·25	2·18	2·10	2·05	2·01	1·96	1·92	1·87	1·81
22	4·30	3·44	3·05	2·82	2·66	2·55	2·46	2·40	2·34	2·30	2·23	2·15	2·07	2·03	1·98	1·94	1·89	1·84	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·44	2·37	2·32	2·27	2·20	2·13	2·05	2·01	1·96	1·91	1·86	1·81	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·42	2·36	2·30	2·25	2·18	2·11	2·03	1·98	1·94	1·89	1·84	1·79	1·73
25	4·24	3·39	2·99	2·76	2·60	2·49	2·40	2·34	2·28	2·24	2·16	2·09	2·01	1·96	1·92	1·87	1·82	1·77	1·71
26	4·23	3·37	2·98	2·74	2·59	2·47	2·39	2·32	2·27	2·22	2·15	2·07	1·99	1·95	1·90	1·85	1·80	1·75	1·69
27	4·21	3·35	2·96	2·73	2·57	2·46	2·37	2·31	2·25	2·20	2·13	2·06	1·97	1·93	1·88	1·84	1·79	1·73	1·67
28	4·20	3·34	2·95	2·71	2·56	2·45	2·36	2·29	2·24	2·19	2·12	2·04	1·96	1·91	1·87	1·82	1·77	1·71	1·65
29	4·18	3·33	2·93	2·70	2·55	2·43	2·35	2·28	2·22	2·18	2·10	2·03	1·94	1·90	1·85	1·81	1·75	1·70	1·64
30	4·17	3·32	2·92	2·69	2·53	2·42	2·33	2·27	2·21	2·16	2·09	2·01	1·93	1·89	1·84	1·79	1·74	1·68	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·25	2·18	2·12	2·08	2·00	1·92	1·84	1·79	1·74	1·69	1·64	1·58	1·51
60	4·00	3·15	2·76	2·53	2·37	2·25	2·17	2·10	2·04	1·99	1·92	1·84	1·75	1·70	1·65	1·59	1·53	1·47	1·39
120	3·92	3·07	2·68	2·45	2·29	2·17	2·09	2·02	1·96	1·91	1·83	1·75	1·66	1·61	1·55	1·50	1·43	1·35	1·25
∞	3·84	3·00	2·60	2·37	2·21	2·10	2·01	1·94	1·88	1·83	1·75	1·67	1·57	1·52	1·46	1·39	1·32	1·22	1·00

$F = \dfrac{s_1^2}{s_2^2} = \dfrac{S_1}{\nu_1}\Big/\dfrac{S_2}{\nu_2}$, where $s_1^2 = S_1/\nu_1$ and $s_2^2 = S_2/\nu_2$ are independent mean squares estimating a common variance σ^2 and based on ν_1 and ν_2 degrees of freedom, respectively.

Percentage Points of the F-distribution (Variance Ratio) (continued)

Upper 2·5 % points

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647·8	799·5	864·2	899·6	921·8	937·1	948·2	956·7	963·3	968·6	976·7	984·9	993·1	997·2	1001	1006	1010	1014	1018
2	38·51	39·00	39·17	39·25	39·30	39·33	39·36	39·37	39·39	39·40	39·41	39·43	39·45	39·46	39·46	39·47	39·48	39·49	39·50
3	17·44	16·04	15·44	15·10	14·88	14·73	14·62	14·54	14·47	14·42	14·34	14·25	14·17	14·12	14·08	14·04	13·99	13·95	13·90
4	12·22	10·65	9·98	9·60	9·36	9·20	9·07	8·98	8·90	8·84	8·75	8·66	8·56	8·51	8·46	8·41	8·36	8·31	8·26
5	10·01	8·43	7·76	7·39	7·15	6·98	6·85	6·76	6·68	6·62	6·52	6·43	6·33	6·28	6·23	6·18	6·12	6·07	6·02
6	8·81	7·26	6·60	6·23	5·99	5·82	5·70	5·60	5·52	5·46	5·37	5·27	5·17	5·12	5·07	5·01	4·96	4·90	4·85
7	8·07	6·54	5·89	5·52	5·29	5·12	4·99	4·90	4·82	4·76	4·67	4·57	4·47	4·42	4·36	4·31	4·25	4·20	4·14
8	7·57	6·06	5·42	5·05	4·82	4·65	4·53	4·43	4·36	4·30	4·20	4·10	4·00	3·95	3·89	3·84	3·78	3·73	3·67
9	7·21	5·71	5·08	4·72	4·48	4·32	4·20	4·10	4·03	3·96	3·87	3·77	3·67	3·61	3·56	3·51	3·45	3·39	3·33
10	6·94	5·46	4·83	4·47	4·24	4·07	3·95	3·85	3·78	3·72	3·62	3·52	3·42	3·37	3·31	3·26	3·20	3·14	3·08
11	6·72	5·26	4·63	4·28	4·04	3·88	3·76	3·66	3·59	3·53	3·43	3·33	3·23	3·17	3·12	3·06	3·00	2·94	2·88
12	6·55	5·10	4·47	4·12	3·89	3·73	3·61	3·51	3·44	3·37	3·28	3·18	3·07	3·02	2·96	2·91	2·85	2·79	2·72
13	6·41	4·97	4·35	4·00	3·77	3·60	3·48	3·39	3·31	3·25	3·15	3·05	2·95	2·89	2·84	2·78	2·72	2·66	2·60
14	6·30	4·86	4·24	3·89	3·66	3·50	3·38	3·29	3·21	3·15	3·05	2·95	2·84	2·79	2·73	2·67	2·61	2·55	2·49
15	6·20	4·77	4·15	3·80	3·58	3·41	3·29	3·20	3·12	3·06	2·96	2·86	2·76	2·70	2·64	2·59	2·52	2·46	2·40
16	6·12	4·69	4·08	3·73	3·50	3·34	3·22	3·12	3·05	2·99	2·89	2·79	2·68	2·63	2·57	2·51	2·45	2·38	2·32
17	6·04	4·62	4·01	3·66	3·44	3·28	3·16	3·06	2·98	2·92	2·82	2·72	2·62	2·56	2·50	2·44	2·38	2·32	2·25
18	5·98	4·56	3·95	3·61	3·38	3·22	3·10	3·01	2·93	2·87	2·77	2·67	2·56	2·50	2·44	2·38	2·32	2·26	2·19
19	5·92	4·51	3·90	3·56	3·33	3·17	3·05	2·96	2·88	2·82	2·72	2·62	2·51	2·45	2·39	2·33	2·27	2·20	2·13
20	5·87	4·46	3·86	3·51	3·29	3·13	3·01	2·91	2·84	2·77	2·68	2·57	2·46	2·41	2·35	2·29	2·22	2·16	2·09
21	5·83	4·42	3·82	3·48	3·25	3·09	2·97	2·87	2·80	2·73	2·64	2·53	2·42	2·37	2·31	2·25	2·18	2·11	2·04
22	5·79	4·38	3·78	3·44	3·22	3·05	2·93	2·84	2·76	2·70	2·60	2·50	2·39	2·33	2·27	2·21	2·14	2·08	2·00
23	5·75	4·35	3·75	3·41	3·18	3·02	2·90	2·81	2·73	2·67	2·57	2·47	2·36	2·30	2·24	2·18	2·11	2·04	1·97
24	5·72	4·32	3·72	3·38	3·15	2·99	2·87	2·78	2·70	2·64	2·54	2·44	2·33	2·27	2·21	2·15	2·08	2·01	1·94
25	5·69	4·29	3·69	3·35	3·13	2·97	2·85	2·75	2·68	2·61	2·51	2·41	2·30	2·24	2·18	2·12	2·05	1·98	1·91
26	5·66	4·27	3·67	3·33	3·10	2·94	2·82	2·73	2·65	2·59	2·49	2·39	2·28	2·22	2·16	2·09	2·03	1·95	1·88
27	5·63	4·24	3·65	3·31	3·08	2·92	2·80	2·71	2·63	2·57	2·47	2·36	2·25	2·19	2·13	2·07	2·00	1·93	1·85
28	5·61	4·22	3·63	3·29	3·06	2·90	2·78	2·69	2·61	2·55	2·45	2·34	2·23	2·17	2·11	2·05	1·98	1·91	1·83
29	5·59	4·20	3·61	3·27	3·04	2·88	2·76	2·67	2·59	2·53	2·43	2·32	2·21	2·15	2·09	2·03	1·96	1·89	1·81
30	5·57	4·18	3·59	3·25	3·03	2·87	2·75	2·65	2·57	2·51	2·41	2·31	2·20	2·14	2·07	2·01	1·94	1·87	1·79
40	5·42	4·05	3·46	3·13	2·90	2·74	2·62	2·53	2·45	2·39	2·29	2·18	2·07	2·01	1·94	1·88	1·80	1·72	1·64
60	5·29	3·93	3·34	3·01	2·79	2·63	2·51	2·41	2·33	2·27	2·17	2·06	1·94	1·88	1·82	1·74	1·67	1·58	1·48
120	5·15	3·80	3·23	2·89	2·67	2·52	2·39	2·30	2·22	2·16	2·05	1·94	1·82	1·76	1·69	1·61	1·53	1·43	1·31
∞	5·02	3·69	3·12	2·79	2·57	2·41	2·29	2·19	2·11	2·05	1·94	1·83	1·71	1·64	1·57	1·48	1·39	1·27	1·00

$F=\dfrac{s_1^2}{s_2^2}=\dfrac{S_1}{\nu_1}\Big/\dfrac{S_2}{\nu_2}$, where $s_1^2=S_1/\nu_1$ and $s_2^2=S_2/\nu_2$ are independent mean squares estimating a common variance σ^2 and based on ν_1 and ν_2 degrees of freedom, respectively.

Percentage Points of the F-distribution (Variance Ratio) (continued)

Upper 1 % points

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999·5	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98·50	99·00	99·17	99·25	99·30	99·33	99·36	99·37	99·39	99·40	99·42	99·43	99·45	99·46	99·47	99·47	99·48	99·49	99·50
3	34·12	30·82	29·46	28·71	28·24	27·91	27·67	27·49	27·35	27·23	27·05	26·87	26·69	26·60	26·50	26·41	26·32	26·22	26·13
4	21·20	18·00	16·69	15·98	15·52	15·21	14·98	14·80	14·66	14·55	14·37	14·20	14·02	13·93	13·84	13·75	13·65	13·56	13·46
5	16·26	13·27	12·06	11·39	10·97	10·67	10·46	10·29	10·16	10·05	9·89	9·72	9·55	9·47	9·38	9·29	9·20	9·11	9·02
6	13·75	10·92	9·78	9·15	8·75	8·47	8·26	8·10	7·98	7·87	7·72	7·56	7·40	7·31	7·23	7·14	7·06	6·97	6·88
7	12·25	9·55	8·45	7·85	7·46	7·19	6·99	6·84	6·72	6·62	6·47	6·31	6·16	6·07	5·99	5·91	5·82	5·74	5·65
8	11·26	8·65	7·59	7·01	6·63	6·37	6·18	6·03	5·91	5·81	5·67	5·52	5·36	5·28	5·20	5·12	5·03	4·95	4·86
9	10·56	8·02	6·99	6·42	6·06	5·80	5·61	5·47	5·35	5·26	5·11	4·96	4·81	4·73	4·65	4·57	4·48	4·40	4·31
10	10·04	7·56	6·55	5·99	5·64	5·39	5·20	5·06	4·94	4·85	4·71	4·56	4·41	4·33	4·25	4·17	4·08	4·00	3·91
11	9·65	7·21	6·22	5·67	5·32	5·07	4·89	4·74	4·63	4·54	4·40	4·25	4·10	4·02	3·94	3·86	3·78	3·69	3·60
12	9·33	6·93	5·95	5·41	5·06	4·82	4·64	4·50	4·39	4·30	4·16	4·01	3·86	3·78	3·70	3·62	3·54	3·45	3·36
13	9·07	6·70	5·74	5·21	4·86	4·62	4·44	4·30	4·19	4·10	3·96	3·82	3·66	3·59	3·51	3·43	3·34	3·25	3·17
14	8·86	6·51	5·56	5·04	4·69	4·46	4·28	4·14	4·03	3·94	3·80	3·66	3·51	3·43	3·35	3·27	3·18	3·09	3·00
15	8·68	6·36	5·42	4·89	4·56	4·32	4·14	4·00	3·89	3·80	3·67	3·52	3·37	3·29	3·21	3·13	3·05	2·96	2·87
16	8·53	6·23	5·29	4·77	4·44	4·20	4·03	3·89	3·78	3·69	3·55	3·41	3·26	3·18	3·10	3·02	2·93	2·84	2·75
17	8·40	6·11	5·18	4·67	4·34	4·10	3·93	3·79	3·68	3·59	3·46	3·31	3·16	3·08	3·00	2·92	2·83	2·75	2·65
18	8·29	6·01	5·09	4·58	4·25	4·01	3·84	3·71	3·60	3·51	3·37	3·23	3·08	3·00	2·92	2·84	2·75	2·66	2·57
19	8·18	5·93	5·01	4·50	4·17	3·94	3·77	3·63	3·52	3·43	3·30	3·15	3·00	2·92	2·84	2·76	2·67	2·58	2·49
20	8·10	5·85	4·94	4·43	4·10	3·87	3·70	3·56	3·46	3·37	3·23	3·09	2·94	2·86	2·78	2·69	2·61	2·52	2·42
21	8·02	5·78	4·87	4·37	4·04	3·81	3·64	3·51	3·40	3·31	3·17	3·03	2·88	2·80	2·72	2·64	2·55	2·46	2·36
22	7·95	5·72	4·82	4·31	3·99	3·76	3·59	3·45	3·35	3·26	3·12	2·98	2·83	2·75	2·67	2·58	2·50	2·40	2·31
23	7·88	5·66	4·76	4·26	3·94	3·71	3·54	3·41	3·30	3·21	3·07	2·93	2·78	2·70	2·62	2·54	2·45	2·35	2·26
24	7·82	5·61	4·72	4·22	3·90	3·67	3·50	3·36	3·26	3·17	3·03	2·89	2·74	2·66	2·58	2·49	2·40	2·31	2·21
25	7·77	5·57	4·68	4·18	3·85	3·63	3·46	3·32	3·22	3·13	2·99	2·85	2·70	2·62	2·54	2·45	2·36	2·27	2·17
26	7·72	5·53	4·64	4·14	3·82	3·59	3·42	3·29	3·18	3·09	2·96	2·81	2·66	2·58	2·50	2·42	2·33	2·23	2·13
27	7·68	5·49	4·60	4·11	3·78	3·56	3·39	3·26	3·15	3·06	2·93	2·78	2·63	2·55	2·47	2·38	2·29	2·20	2·10
28	7·64	5·45	4·57	4·07	3·75	3·53	3·36	3·23	3·12	3·03	2·90	2·75	2·60	2·52	2·44	2·35	2·26	2·17	2·06
29	7·60	5·42	4·54	4·04	3·73	3·50	3·33	3·20	3·09	3·00	2·87	2·73	2·57	2·49	2·41	2·33	2·23	2·14	2·03
30	7·56	5·39	4·51	4·02	3·70	3·47	3·30	3·17	3·07	2·98	2·84	2·70	2·55	2·47	2·39	2·30	2·21	2·11	2·01
40	7·31	5·18	4·31	3·83	3·51	3·29	3·12	2·99	2·89	2·80	2·66	2·52	2·37	2·29	2·20	2·11	2·02	1·92	1·80
60	7·08	4·98	4·13	3·65	3·34	3·12	2·95	2·82	2·72	2·63	2·50	2·35	2·20	2·12	2·03	1·94	1·84	1·73	1·60
120	6·85	4·79	3·95	3·48	3·17	2·96	2·79	2·66	2·56	2·47	2·34	2·19	2·03	1·95	1·86	1·76	1·66	1·53	1·38
∞	6·63	4·61	3·78	3·32	3·02	2·80	2·64	2·51	2·41	2·32	2·18	2·04	1·88	1·79	1·70	1·59	1·47	1·32	1·00

$F = \dfrac{s_1^2}{s_2^2} = \dfrac{S_1}{\nu_1}\Big/\dfrac{S_2}{\nu_2}$, where $s_1^2 = S_1/\nu_1$ and $s_2^2 = S_2/\nu_2$ are independent mean squares estimating a common variance σ^2 and based on ν_1 and ν_2 degrees of freedom, respectively.

Percentage Points of the F-distribution (Variance Ratio) (continued)

Upper 0·5 % points

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198·5	199·0	199·2	199·2	199·3	199·3	199·4	199·4	199·4	199·4	199·4	199·4	199·4	199·5	199·5	199·5	199·5	199·5	199·5
3	55·55	49·80	47·47	46·19	45·39	44·84	44·43	44·13	43·88	43·69	43·39	43·08	42·78	42·62	42·47	42·31	42·15	41·99	41·83
4	31·33	26·28	24·26	23·15	22·46	21·97	21·62	21·35	21·14	20·97	20·70	20·44	20·17	20·03	19·89	19·75	19·61	19·47	19·32
5	22·78	18·31	16·53	15·56	14·94	14·51	14·20	13·96	13·77	13·62	13·38	13·15	12·90	12·78	12·66	12·53	12·40	12·27	12·14
6	18·63	14·54	12·92	12·03	11·46	11·07	10·79	10·57	10·39	10·25	10·03	9·81	9·59	9·47	9·36	9·24	9·12	9·00	8·88
7	16·24	12·40	10·88	10·05	9·52	9·16	8·89	8·68	8·51	8·38	8·18	7·97	7·75	7·65	7·53	7·42	7·31	7·19	7·08
8	14·69	11·04	9·60	8·81	8·30	7·95	7·69	7·50	7·34	7·21	7·01	6·81	6·61	6·50	6·40	6·29	6·18	6·06	5·95
9	13·61	10·11	8·72	7·96	7·47	7·13	6·88	6·69	6·54	6·42	6·23	6·03	5·83	5·73	5·62	5·52	5·41	5·30	5·19
10	12·83	9·43	8·08	7·34	6·87	6·54	6·30	6·12	5·97	5·85	5·66	5·47	5·27	5·17	5·07	4·97	4·86	4·75	4·64
11	12·23	8·91	7·60	6·88	6·42	6·10	5·86	5·68	5·54	5·42	5·24	5·05	4·86	4·76	4·65	4·55	4·44	4·34	4·23
12	11·75	8·51	7·23	6·52	6·07	5·76	5·52	5·35	5·20	5·09	4·91	4·72	4·53	4·43	4·33	4·23	4·12	4·01	3·90
13	11·37	8·19	6·93	6·23	5·79	5·48	5·25	5·08	4·94	4·82	4·64	4·46	4·27	4·17	4·07	3·97	3·87	3·76	3·65
14	11·06	7·92	6·68	6·00	5·56	5·26	5·03	4·86	4·72	4·60	4·43	4·25	4·06	3·96	3·86	3·76	3·66	3·55	3·44
15	10·80	7·70	6·48	5·80	5·37	5·07	4·85	4·67	4·54	4·42	4·25	4·07	3·88	3·79	3·69	3·58	3·48	3·37	3·26
16	10·58	7·51	6·30	5·64	5·21	4·91	4·69	4·52	4·38	4·27	4·10	3·92	3·73	3·64	3·54	3·44	3·33	3·22	3·11
17	10·38	7·35	6·16	5·50	5·07	4·78	4·56	4·39	4·25	4·14	3·97	3·79	3·61	3·51	3·41	3·31	3·21	3·10	2·98
18	10·22	7·21	6·03	5·37	4·96	4·66	4·44	4·28	4·14	4·03	3·86	3·68	3·50	3·40	3·30	3·20	3·10	2·99	2·87
19	10·07	7·09	5·92	5·27	4·85	4·56	4·34	4·18	4·04	3·93	3·76	3·59	3·40	3·31	3·21	3·11	3·00	2·89	2·78
20	9·94	6·99	5·82	5·17	4·76	4·47	4·26	4·09	3·96	3·85	3·68	3·50	3·32	3·22	3·12	3·02	2·92	2·81	2·69
21	9·83	6·89	5·73	5·09	4·68	4·39	4·18	4·01	3·88	3·77	3·60	3·43	3·24	3·15	3·05	2·95	2·84	2·73	2·61
22	9·73	6·81	5·65	5·02	4·61	4·32	4·11	3·94	3·81	3·70	3·54	3·36	3·18	3·08	2·98	2·88	2·77	2·66	2·55
23	9·63	6·73	5·58	4·95	4·54	4·26	4·05	3·88	3·75	3·64	3·47	3·30	3·12	3·02	2·92	2·82	2·71	2·60	2·48
24	9·55	6·66	5·52	4·89	4·49	4·20	3·99	3·83	3·69	3·59	3·42	3·25	3·06	2·97	2·87	2·77	2·66	2·55	2·43
25	9·48	6·60	5·46	4·84	4·43	4·15	3·94	3·78	3·64	3·54	3·37	3·20	3·01	2·92	2·82	2·72	2·61	2·50	2·38
26	9·41	6·54	5·41	4·79	4·38	4·10	3·89	3·73	3·60	3·49	3·33	3·15	2·97	2·87	2·77	2·67	2·56	2·45	2·33
27	9·34	6·49	5·36	4·74	4·34	4·06	3·85	3·69	3·56	3·45	3·28	3·11	2·93	2·83	2·73	2·63	2·52	2·41	2·29
28	9·28	6·44	5·32	4·70	4·30	4·02	3·81	3·65	3·52	3·41	3·25	3·07	2·89	2·79	2·69	2·59	2·48	2·37	2·25
29	9·23	6·40	5·28	4·66	4·26	3·98	3·77	3·61	3·48	3·38	3·21	3·04	2·86	2·76	2·66	2·56	2·45	2·33	2·21
30	9·18	6·35	5·24	4·62	4·23	3·95	3·74	3·58	3·45	3·34	3·18	3·01	2·82	2·73	2·63	2·52	2·42	2·30	2·18
40	8·83	6·07	4·98	4·37	3·99	3·71	3·51	3·35	3·22	3·12	2·95	2·78	2·60	2·50	2·40	2·30	2·18	2·06	1·93
60	8·49	5·79	4·73	4·14	3·76	3·49	3·29	3·13	3·01	2·90	2·74	2·57	2·39	2·29	2·19	2·08	1·96	1·83	1·69
120	8·18	5·54	4·50	3·92	3·55	3·28	3·09	2·93	2·81	2·71	2·54	2·37	2·19	2·09	1·98	1·87	1·75	1·61	1·43
∞	7·88	5·30	4·28	3·72	3·35	3·09	2·90	2·74	2·62	2·52	2·36	2·19	2·00	1·90	1·79	1·67	1·53	1·36	1·00

$F = \dfrac{s_1^2}{s_2^2} = \dfrac{S_1}{\nu_1}\Big/\dfrac{S_2}{\nu_2}$, where $s_1^2 = S_1/\nu_1$ and $s_2^2 = S_2/\nu_2$ are independent mean squares estimating a common variance σ^2 and based on ν_1 and ν_2 degrees of freedom, respectively

Percentage Points of the F-distribution (Variance Ratio) (continued)

Upper 0·1 % points

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998·5	999·0	999·2	999·2	999·3	999·3	999·4	999·4	999·4	999·4	999·4	999·4	999·4	999·5	999·5	999·5	999·5	999·5	999·5
3	167·0	148·5	141·1	137·1	134·6	132·8	131·6	130·6	129·9	129·2	128·3	127·4	126·4	125·9	125·4	125·0	124·5	124·0	123·5
4	74·14	61·25	56·18	53·44	51·71	50·53	49·66	49·00	48·47	48·05	47·41	46·76	46·10	45·77	45·43	45·09	44·75	44·40	44·05
5	47·18	37·12	33·20	31·09	29·75	28·84	28·16	27·64	27·24	26·92	26·42	25·91	25·39	25·14	24·87	24·60	24·33	24·06	23·79
6	35·51	27·00	23·70	21·92	20·81	20·03	19·46	19·03	18·69	18·41	17·99	17·56	17·12	16·89	16·67	16·44	16·21	15·99	15·75
7	29·25	21·69	18·77	17·19	16·21	15·52	15·02	14·63	14·33	14·08	13·71	13·32	12·93	12·73	12·53	12·33	12·12	11·91	11·70
8	25·42	18·49	15·83	14·39	13·49	12·86	12·40	12·04	11·77	11·54	11·19	10·84	10·48	10·30	10·11	9·92	9·73	9·53	9·33
9	22·86	16·39	13·90	12·56	11·71	11·13	10·70	10·37	10·11	9·89	9·57	9·24	8·90	8·72	8·55	8·37	8·19	8·00	7·81
10	21·04	14·91	12·55	11·28	10·48	9·92	9·52	9·20	8·96	8·75	8·45	8·13	7·80	7·64	7·47	7·30	7·12	6·94	6·76
11	19·69	13·81	11·56	10·35	9·58	9·05	8·66	8·35	8·12	7·92	7·63	7·32	7·01	6·85	6·68	6·52	6·35	6·17	6·00
12	18·64	12·97	10·80	9·63	8·89	8·38	8·00	7·71	7·48	7·29	7·00	6·71	6·40	6·25	6·09	5·93	5·76	5·59	5·42
13	17·81	12·31	10·21	9·07	8·35	7·86	7·49	7·21	6·98	6·80	6·52	6·23	5·93	5·78	5·63	5·47	5·30	5·14	4·97
14	17·14	11·78	9·73	8·62	7·92	7·43	7·08	6·80	6·58	6·40	6·13	5·85	5·56	5·41	5·25	5·10	4·94	4·77	4·60
15	16·59	11·34	9·34	8·25	7·57	7·09	6·74	6·47	6·26	6·08	5·81	5·54	5·25	5·10	4·95	4·80	4·64	4·47	4·31
16	16·12	10·97	9·00	7·94	7·27	6·81	6·46	6·19	5·98	5·81	5·55	5·27	4·99	4·85	4·70	4·54	4·39	4·23	4·06
17	15·72	10·66	8·73	7·68	7·02	6·56	6·22	5·96	5·75	5·58	5·32	5·05	4·78	4·63	4·48	4·33	4·18	4·02	3·85
18	15·38	10·39	8·49	7·46	6·81	6·35	6·02	5·76	5·56	5·39	5·13	4·87	4·59	4·45	4·30	4·15	4·00	3·84	3·67
19	15·08	10·16	8·28	7·26	6·62	6·18	5·85	5·59	5·39	5·22	4·97	4·70	4·43	4·29	4·14	3·99	3·84	3·68	3·51
20	14·82	9·95	8·10	7·10	6·46	6·02	5·69	5·44	5·24	5·08	4·82	4·56	4·29	4·15	4·00	3·86	3·70	3·54	3·38
21	14·59	9·77	7·94	6·95	6·32	5·88	5·56	5·31	5·11	4·95	4·70	4·44	4·17	4·03	3·88	3·74	3·58	3·42	3·26
22	14·38	9·61	7·80	6·81	6·19	5·76	5·44	5·19	4·99	4·83	4·58	4·33	4·06	3·92	3·78	3·63	3·48	3·32	3·15
23	14·19	9·47	7·67	6·69	6·08	5·65	5·33	5·09	4·89	4·73	4·48	4·23	3·96	3·82	3·68	3·53	3·38	3·22	3·05
24	14·03	9·34	7·55	6·59	5·98	5·55	5·23	4·99	4·80	4·64	4·39	4·14	3·87	3·74	3·59	3·45	3·29	3·14	2·97
25	13·88	9·22	7·45	6·49	5·88	5·46	5·15	4·91	4·71	4·56	4·31	4·06	3·79	3·66	3·52	3·37	3·22	3·06	2·89
26	13·74	9·12	7·36	6·41	5·80	5·38	5·07	4·83	4·64	4·48	4·24	3·99	3·72	3·59	3·44	3·30	3·15	2·99	2·82
27	13·61	9·02	7·27	6·33	5·73	5·31	5·00	4·76	4·57	4·41	4·17	3·92	3·66	3·52	3·38	3·23	3·08	2·92	2·75
28	13·50	8·93	7·19	6·25	5·66	5·24	4·93	4·69	4·50	4·35	4·11	3·86	3·60	3·46	3·32	3·18	3·02	2·86	2·69
29	13·39	8·85	7·12	6·19	5·59	5·18	4·87	4·64	4·45	4·29	4·05	3·80	3·54	3·41	3·27	3·12	2·97	2·81	2·64
30	13·29	8·77	7·05	6·12	5·53	5·12	4·82	4·58	4·39	4·24	4·00	3·75	3·49	3·36	3·22	3·07	2·92	2·76	2·59
40	12·61	8·25	6·60	5·70	5·13	4·73	4·44	4·21	4·02	3·87	3·64	3·40	3·15	3·01	2·87	2·73	2·57	2·41	2·23
60	11·97	7·76	6·17	5·31	4·76	4·37	4·09	3·87	3·69	3·54	3·31	3·08	2·83	2·69	2·55	2·41	2·25	2·08	1·89
120	11·38	7·32	5·79	4·95	4·42	4·04	3·77	3·55	3·38	3·24	3·02	2·78	2·53	2·40	2·26	2·11	1·95	1·76	1·54
∞	10·83	6·91	5·42	4·62	4·10	3·74	3·47	3·27	3·10	2·96	2·74	2·51	2·27	2·13	1·99	1·84	1·66	1·45	1·00

* Multiply these entries by 100.

This 0·1 % table is based on the following sources: Colcord & Deming (1935); Fisher & Yates (1953, Table V) used with the permission of the authors and of Messrs Oliver and Boyd; Norton (1952).

This table was reprinted from Biometrika Tables for Statisticians, Vol. 1, 3rd Edition, Table 18, with the permission of the Biometrika Trustees.

Percentage Points of the t-distribution

ν	$Q=0{\cdot}4$ $2Q=0{\cdot}8$	0·25 0·5	0·05 0·1	0·025 0·05	0·005 0·01	0·0025 0·005	0·0005 0·001
1	0·325	1·000	6·314	12·706	63·657	127·32	636·62
2	·289	0·816	2·920	4·303	9·925	14·089	31·598
3	·277	·765	2·353	3·182	5·841	7·453	12·924
4	·271	·741	2·132	2·776	4·604	5·598	8·610
5	0·267	0·727	2·015	2·571	4·032	4·773	6·869
6	·265	·718	1·943	2·447	3·707	4·317	5·959
7	·263	·711	1·895	2·365	3·499	4·029	5·408
8	·262	·706	1·860	2·306	3·355	3·833	5·041
9	·261	·703	1·833	2·262	3·250	3·690	4·781
10	0·260	0·700	1·812	2·228	3·169	3·581	4·587
11	·260	·697	1·796	2·201	3·106	3·497	4·437
12	·259	·695	1·782	2·179	3·055	3·428	4·318
13	·259	·694	1·771	2·160	3·012	3·372	4·221
14	·258	·692	1·761	2·145	2·977	3·326	4·140
15	0·258	0·691	1·753	2·131	2·947	3·286	4·073
16	·258	·690	1·746	2·120	2·921	3·252	4·015
17	·257	·689	1·740	2·110	2·898	3·222	3·965
18	·257	·688	1·734	2·101	2·878	3·197	3·922
19	·257	·688	1·729	2·093	2·861	3·174	3·883
20	0·257	0·687	1·725	2·086	2·845	3·153	3·850
21	·257	·686	1·721	2·080	2·831	3·135	3·819
22	·256	·686	1·717	2·074	2·819	3·119	3·792
23	·256	·685	1·714	2·069	2·807	3·104	3·767
24	·256	·685	1·711	2·064	2·797	3·091	3·745
25	0·256	0·684	1·708	2·060	2·787	3·078	3·725
26	·256	·684	1·706	2·056	2·779	3·067	3·707
27	·256	·684	1·703	2·052	2·771	3·057	3·690
28	·256	·683	1·701	2·048	2·763	3·047	3·674
29	·256	·683	1·699	2·045	2·756	3·038	3·659
30	0·256	0·683	1·697	2·042	2·750	3·030	3·646
40	·255	·681	1·684	2·021	2·704	2·971	3·551
60	·254	·679	1·671	2·000	2·660	2·915	3·460
120	·254	·677	1·658	1·980	2·617	2·860	3·373
∞	·253	·674	1·645	1·960	2·576	2·807	3·291

$Q = 1 - P(t|\nu)$ is the upper-tail area of the distribution for ν degrees of freedom, appropriate for use in a single-tail test. For a two-tail test, $2Q$ must be used.

This table was reprinted from Biometrika Tables for Statisticians, Vol. 1, 3rd Edition, Table 12, with the permission of the Biometrika Trustees.

Percentage Points of the X^2-Distribution

ν \ Q	0·995	0·990	0·975	0·950	0·900	0·750	0·500
1	392704.10^{-10}	157088.10^{-9}	982069.10^{-9}	393214.10^{-8}	0·0157908	0·1015308	0·454936
2	0·0100251	0·0201007	0·0506356	0·102587	0·210721	0·575364	1·38629
3	0·0717218	0·114832	0·215795	0·351846	0·584374	1·212534	2·36597
4	0·206989	0·297109	0·484419	0·710723	1·063623	1·92256	3·35669
5	0·411742	0·554298	0·831212	1·145476	1·61031	2·67460	4·35146
6	0·675727	0·872090	1·23734	1·63538	2·20413	3·45460	5·34812
7	0·989256	1·239043	1·68987	2·16735	2·83311	4·25485	6·34581
8	1·34441	1·64650	2·17973	2·73264	3·48954	5·07064	7·34412
9	1·73493	2·08790	2·70039	3·32511	4·16816	5·89883	8·34283
10	2·15586	2·55821	3·24697	3·94030	4·86518	6·73720	9·34182
11	2·60322	3·05348	3·81575	4·57481	5·57778	7·58414	10·3410
12	3·07382	3·57057	4·40379	5·22603	6·30380	8·43842	11·3403
13	3·56503	4·10692	5·00875	5·89186	7·04150	9·29907	12·3398
14	4·07467	4·66043	5·62873	6·57063	7·78953	10·1653	13·3393
15	4·60092	5·22935	6·26214	7·26094	8·54676	11·0365	14·3389
16	5·14221	5·81221	6·90766	7·96165	9·31224	11·9122	15·3385
17	5·69722	6·40776	7·56419	8·67176	10·0852	12·7919	16·3382
18	6·26480	7·01491	8·23075	9·39046	10·8649	13·6753	17·3379
19	6·84397	7·63273	8·90652	10·1170	11·6509	14·5620	18·3377
20	7·43384	8·26040	9·59078	10·8508	12·4426	15·4518	19·3374
21	8·03365	8·89720	10·28293	11·5913	13·2396	16·3444	20·3372
22	8·64272	9·54249	10·9823	12·3380	14·0415	17·2396	21·3370
23	9·26043	10·19567	11·6886	13·0905	14·8480	18·1373	22·3369
24	9·88623	10·8564	12·4012	13·8484	15·6587	19·0373	23·3367
25	10·5197	11·5240	13·1197	14·6114	16·4734	19·9393	24·3366
26	11·1602	12·1981	13·8439	15·3792	17·2919	20·8434	25·3365
27	11·8076	12·8785	14·5734	16·1514	18·1139	21·7494	26·3363
28	12·4613	13·5647	15·3079	16·9279	18·9392	22·6572	27·3362
29	13·1211	14·2565	16·0471	17·7084	19·7677	23·5666	28·3361
30	13·7867	14·9535	16·7908	18·4927	20·5992	24·4776	29·3360
40	20·7065	22·1643	24·4330	26·5093	29·0505	33·6603	39·3353
50	27·9907	29·7067	32·3574	34·7643	37·6886	42·9421	49·3349
60	35·5345	37·4849	40·4817	43·1880	46·4589	52·2938	59·3347
70	43·2752	45·4417	48·7576	51·7393	55·3289	61·6983	69·3345
80	51·1719	53·5401	57·1532	60·3915	64·2778	71·1445	79·3343
90	59·1963	61·7541	65·6466	69·1260	73·2911	80·6247	89·3342
100	67·3276	70·0649	74·2219	77·9295	82·3581	90·1332	99·3341
X	−2·5758	−2·3263	−1·9600	−1·6449	−1·2816	−0·6745	0·0000

$$Q = Q(\chi^2 \mid \nu) = 1 - P(\chi^2 \mid \nu) = 2^{-\frac{1}{2}\nu}\{\Gamma(\tfrac{1}{2}\nu)\}^{-1}\int_{\chi^2}^{\infty} e^{-\frac{1}{2}x}\, x^{\frac{1}{2}\nu - 1}\, dx.$$

Percentage Points of the χ^2-Distribution (continued)

ν \ Q	0·250	0·100	0·050	0·025	0·010	0·005	0·001
1	1·32330	2·70554	3·84146	5·02389	6·63490	7·87944	10·828
2	2·77259	4·60517	5·99146	7·37776	9·21034	10·5966	13·816
3	4·10834	6·25139	7·81473	9·34840	11·3449	12·8382	16·266
4	5·38527	7·77944	9·48773	11·1433	13·2767	14·8603	18·467
5	6·62568	9·23636	11·0705	12·8325	15·0863	16·7496	20·515
6	7·84080	10·6446	12·5916	14·4494	16·8119	18·5476	22·458
7	9·03715	12·0170	14·0671	16·0128	18·4753	20·2777	24·322
8	10·2189	13·3616	15·5073	17·5345	20·0902	21·9550	26·125
9	11·3888	14·6837	16·9190	19·0228	21·6660	23·5894	27·877
10	12·5489	15·9872	18·3070	20·4832	23·2093	25·1882	29·588
11	13·7007	17·2750	19·6751	21·9200	24·7250	26·7568	31·264
12	14·8454	18·5493	21·0261	23·3367	26·2170	28·2995	32·909
13	15·9839	19·8119	22·3620	24·7356	27·6882	29·8195	34·528
14	17·1169	21·0641	23·6848	26·1189	29·1412	31·3194	36·123
15	18·2451	22·3071	24·9958	27·4884	30·5779	32·8013	37·697
16	19·3689	23·5418	26·2962	28·8454	31·9999	34·2672	39·252
17	20·4887	24·7690	27·5871	30·1910	33·4087	35·7185	40·790
18	21·6049	25·9894	28·8693	31·5264	34·8053	37·1565	42·312
19	22·7178	27·2036	30·1435	32·8523	36·1909	38·5823	43·820
20	23·8277	28·4120	31·4104	34·1696	37·5662	39·9968	45·315
21	24·9348	29·6151	32·6706	35·4789	38·9322	41·4011	46·797
22	26·0393	30·8133	33·9244	36·7807	40·2894	42·7957	48·268
23	27·1413	32·0069	35·1725	38·0756	41·6384	44·1813	49·728
24	28·2412	33·1962	36·4150	39·3641	42·9798	45·5585	51·179
25	29·3389	34·3816	37·6525	40·6465	44·3141	46·9279	52·618
26	30·4346	35·5632	38·8851	41·9232	45·6417	48·2899	54·052
27	31·5284	36·7412	40·1133	43·1945	46·9629	49·6449	55·476
28	32·6205	37·9159	41·3371	44·4608	48·2782	50·9934	56·892
29	33·7109	39·0875	42·5570	45·7223	49·5879	52·3356	58·301
30	34·7997	40·2560	43·7730	46·9792	50·8922	53·6720	59·703
40	45·6160	51·8051	55·7585	59·3417	63·6907	66·7660	73·402
50	56·3336	63·1671	67·5048	71·4202	76·1539	79·4900	86·661
60	66·9815	74·3970	79·0819	83·2977	88·3794	91·9517	99·607
70	77·5767	85·5270	90·5312	95·0232	100·425	104·215	112·317
80	88·1303	96·5782	101·879	106·629	112·329	116·321	124·839
90	98·6499	107·565	113·145	118·136	124·116	128·299	137·208
100	109·141	118·498	124·342	129·561	135·807	140·169	149·449
X	+0·6745	+1·2816	+1·6449	+1·9600	+2·3263	+2·5758	+3·0902

For $\nu > 100$ take

$$\chi^2 = \nu\left\{1 - \frac{2}{9\nu} + X\sqrt{\frac{2}{9\nu}}\right\}^3 \quad \text{or} \quad \chi^2 = \tfrac{1}{2}\{X + \sqrt{(2\nu - 1)}\}^2,$$

according to the degree of accuracy required. X is the standardized normal deviate corresponding to $P = 1 - Q$, and is shown in the bottom line of the table.

This table was reprinted from Biometrika Tables for Statisticians, Vol. 1, 3rd Edition, Table 8, with the permission of the Biometrika Trustees.

Notes

HEWLETT PACKARD

Part No. 98820-13111
E0782

Printed in U.S.A.
First Edition; July 1982